STAT 244
Lecture 1
9/27/16

Broadly speaking, we will cover 4 topics:

I. Probability

II. Bayesian Inference

III. Maximum Likelihood Inference

IV. Hypothesis Testing

See the syllabus on chalk for organizational details.

I. Probability

Probability Theory

Rules for calculating some probabilities from others, going from simple situations, to complex. [The most simple situation is one where each "outcome" is equally likely.

Definitions:

Given an "experiment" (some process of observation):

Sample Space S = set of all possible outcomes

Event (e.g. E) = a set of outcomes

With events E and F:

E^c = the complement of E ("not E")

$E \cap F$ = the intersection of E and F ("both")

$E \cup F$ = the union of E and F ("E or F or both")

E and F are mutually exclusive if $E \cap F = \emptyset$ ("empty")

Example: $(A \cap B^c) \cup (A \cap B) = A$

so $P((A \cap B^c) \cup (A \cap B)) = P(A)$ (whatever "P" means!)

Probability is a measure.

A measure of what??

uncertainty

belief

relative frequency in the "long run"

<u>Properties:</u>

For any probability measure:

$$P(S) = 1; \quad 0 \leq P(E) \text{ for all } E$$

$$P(E \cup F) = P(E) + P(F) \text{ } \underline{\text{if}} \text{ } E, F \text{ mutually exclusive.}$$

Hence, since $S = E \cup E^c$, $P(E) + P(E^c) = 1$

also, for any E, F:

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

<u>Countable additivity:</u>

If $A_1, A_2, \ldots$ are mutually exclusive,

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \cdots$$

permutations

Permutations and combinations

$P_{r,n}$ = number of ways to choose r objects from n distinguishable where order makes a difference

$$P_{r,n} = \frac{n!}{(n-r)!} = \frac{n(n-1)(n-2)\cdots 2\cdot 1}{(n-r)(n-r-1)\cdots 2\cdot 1}$$

combinations

$C_{r,n}$ = number of ways of choosing r objects from n distinguishable objects where order does <u>not</u> make a difference:

$$C_{r,n} = \binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{n(n-1)(n-2)\cdots 2\cdot 1}{r(r-1)\cdots 2\cdot 1\cdot (n-r)(n-r-1)\cdots 1}$$

$(0! = 1)$

Example: If $r=2$ people are chosen from $n=5$ people to be designated president and vice president, there are $P_{2,5} = 20$ ways to make the selection. If they are to make a committee of two equals, so (A, B) and (B, A) are the same committee, then there are only $\binom{5}{2} = 10$ ways to select

Conditional Probability

(sometimes called "relative probability")

Definition: If $P(F) > 0$,

$$P(E|F) = \frac{P(E \wedge F)}{P(F)}$$

Note: <u>All</u> probs are conditional, because $P(E) = P(E|S)$

also: note that

$$P(E \wedge F) = P(E|F)\,P(F)$$

even if $P(F) = 0$.

<u>Any</u> two of these determine the third.

If $P(E|F) = P(E)$, then E and F are <u>independent events</u>.

Equivalently, $P(E|F) = P(E|F^c)$

<u>or</u> $P(E \wedge F) = P(E)P(F)$

<u>Example</u>: Monte Hall Game

Three doors (A, B, C). One prize. You pick A. Host shows B (empty). Should you switch?

Monte Hall Game

Three doors (A, B, C), one prize
You pick A, host shows B (empty).
Should you swith your guess to C? Assume you do.

S has 6 outcomes:

Prize in A, you see B Prob $= 1/6$
Prize in A, you see C Prob $= 1/6$
Prize in B, you see B Prob $= 0$
Prize in B, you see C Prob $= 1/3$
Prize in C, you see C Prob $= 0$
Prize in C, you see B Prob $= \frac{1}{3}$

$$P(\text{Win} \mid \text{See } B) = \frac{P(\text{Win AND See } B)}{P(\text{See } B)}$$

$$= \frac{\left(\frac{1}{3}\right)}{\left(\frac{1}{3} + \frac{1}{6}\right)} = \frac{2}{3}$$

Random Variables —

functions that take on numerical values at each point in the sample space.

eg — flip two coins, so

$S = \{HH, HT, TH, TT\}$

Let X be number of heads, so

$X = x = 0, 1, 2$

A **Probability Distribution**

$p_X(x)$ (or $p(x)$) is a list of possible values of X and their probabilities.

For the case above

X	0	1	2
$p(x)$	1/4	1/2	1/4

Discrete random varibles: countable number of values

Continuous random variables: values form an interval

Cumulative Distribution Function (cdf)

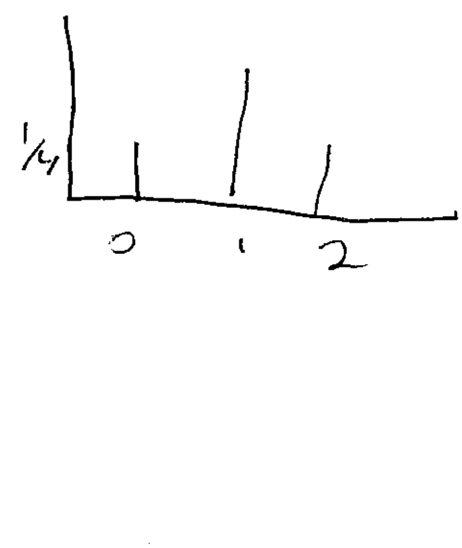
$$F_X(x) \text{ or } F(x) = P(X \le x)$$

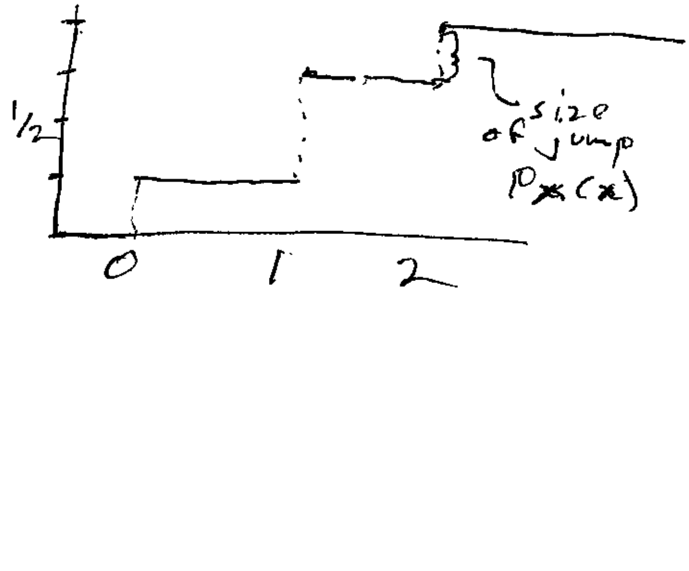
$$= \sum_{a \le x} p(a)$$

Example: (same two coins)

X	0	1	2
$p_X(x)$	1/4	1/2	1/4
$F_X(x)$	1/4	3/4	1

$p_X(x)$





Some Important Examples

1. (very simple) Bernoulli random var: $S = \{0, 1\}$ $p(1) = p$ $p(0) = 1-p$

2. Binomial random var.

This describes the Binomial Experiment:

n independent trials

$\theta = P(\text{"success"})$ each trial... ie each trial is described by an iid (independently identically distributed) Bernoulli random var

X = number of successes

Example: Flip 3 coins, count heads.

$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

$P(X=2) = P(HHT) + P(HTH) + P(THH)$

$P(HHT) = \theta\theta(1-\theta) = P(HTH) = P(THH)$

$\Rightarrow P(X=2) = 3\theta^2(1-\theta)$

In general:

n idep trials, A = "success"
A^c = "failure"

→ in each trial $P(A) = \theta$

X is # of A's

S has 2^n points. For a point in S with x A's and $n-x$ A^c's,

$$P(\underbrace{AA\ldots A}_{x}\underbrace{A^cA^c\ldots A^c}_{n-x}) = \theta^x(1-\theta)^{n-x}$$

$X \in \{0, 1, \ldots, n\}$, a binomial random var.

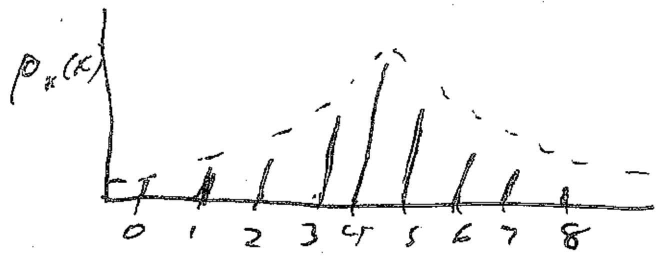
We need to count the number of ways x successes can be chosen from n trials... but that is just $\binom{n}{x}$. The probability of x success is then

$$b(x; n, \theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

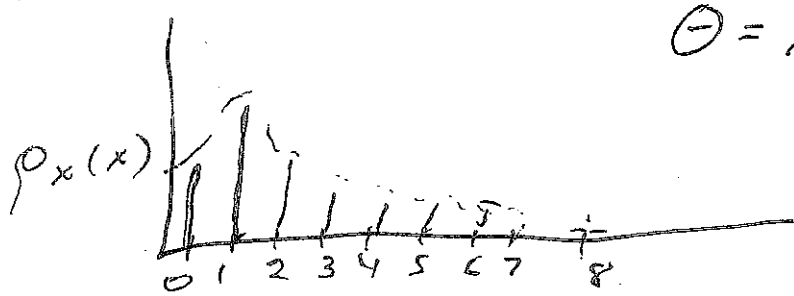
The Binomial Distribution

$\theta = \frac{1}{2}$

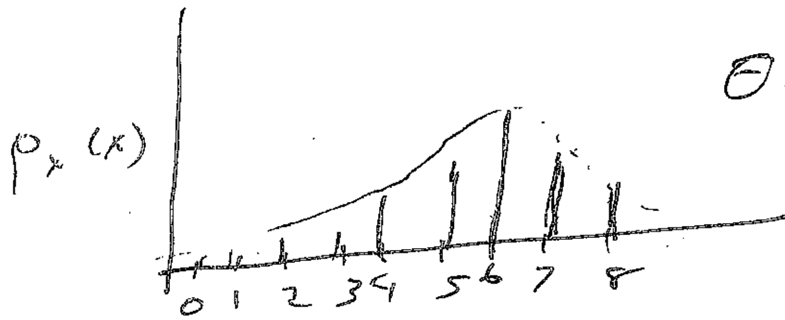
$n = 8$



$\theta = 1/8$



$\theta = 3/4$



$\text{Bin}(n, \theta)$ n and θ are <u>parameters</u>

Review: The Binomial Distribution

STAT 24400 Lecture 2

In general:

n idep trials, $A =$ "success"

$A^c =$ "failure"

→ in each trial $P(A) = \theta$

X is # of A's

S has 2^n points. For a point in S with x A's and $n-x$ A^c's,

$$P(\underbrace{AA\ldots A}_{x}\underbrace{A^cA^c\ldots A^c}_{n-x}) = \theta^x(1-\theta)^{n-x}$$

$X \in \{0, 1, \ldots, n\}$, a binomial random var.

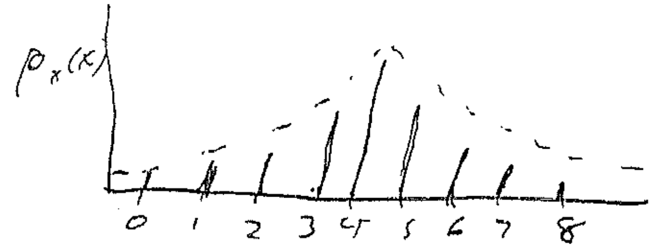
We need to count the number of ways x successes can be chosen from n trials... but that is just $\binom{n}{x}$. The probability of x success is then

$$b(x; n, \theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

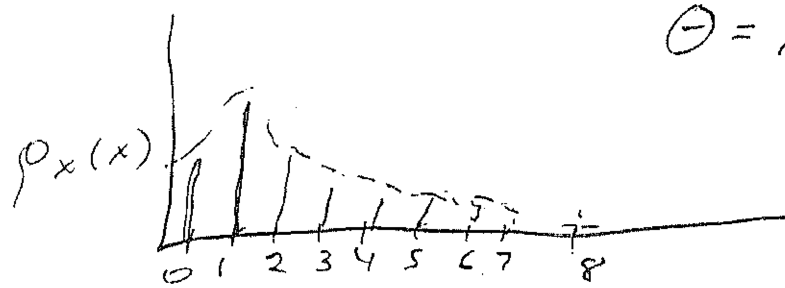
The Binomial Distribution

$\theta = \frac{1}{2}$

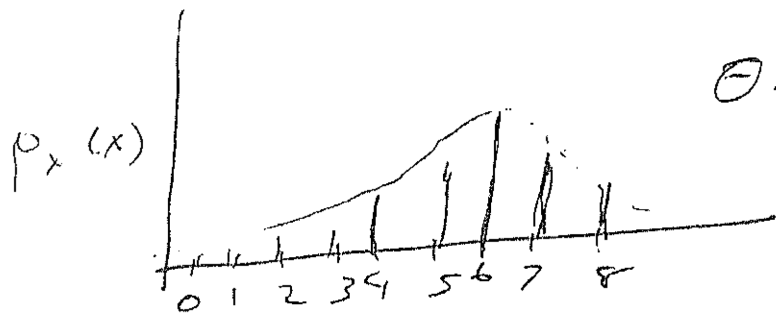
$n = 8$



$\theta = 1/8$



$\theta = 3/4$



$\text{Bin}(n, \theta)$ n and θ are <u>parameters</u>. specifying them determines the distribution.

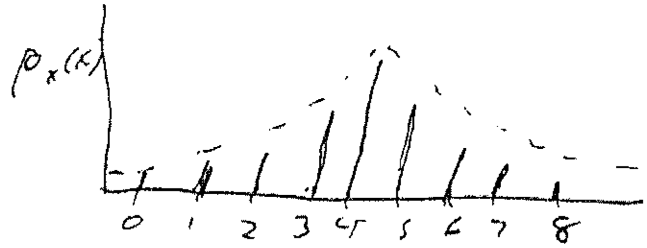
The Negative Binomial Distribution

Perform Bernoulli trials with prob of success θ until there are r successes and X failures

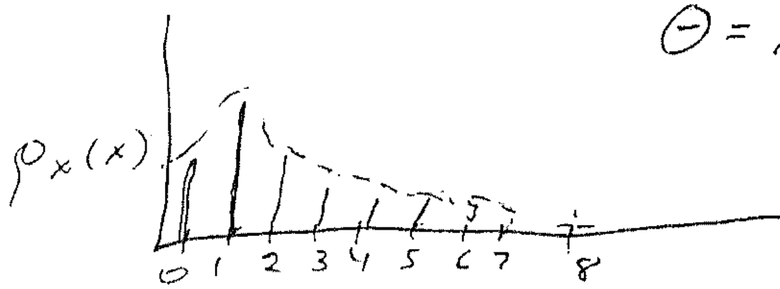
The Binomial Distribution

$\theta = \frac{1}{2}$

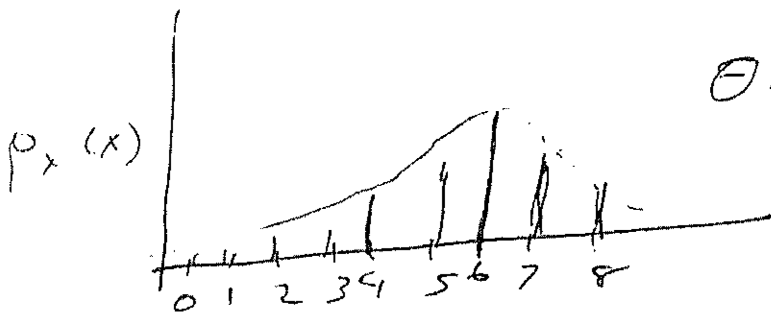
$n = 8$



$\theta = 1/8$



$\theta = 3/4$



$Bin(n, \theta)$ n and θ are <u>parameters</u>. specifying them determines the distribution.

The Negative Binomial Distribution

Perform Bernoulli trials with prob of success θ until there are r successes and X failures

(Neg Binomial, continued)

Example for $r=1$ (this case has a special name: the "geometric distribution")

S	X	Prob
A	0	θ
A^cA	1	$\theta(1-\theta)$
A^cA^cA	2	$\theta(1-\theta)^2$
$A^cA^cA^cA$	3	$\theta(1-\theta)^3$
$\vdots$	$\vdots$	$\vdots$
	k	$\theta(1-\theta)^k$

Negative Binomial, in general:

probability of r successes: θ^r

" $\underbrace{\text{of } k \text{ failures before the } r^{th}}_{\text{each such outcome}}$ success: $(1-\theta)^k$

To find the prob that $X=k$, we

1. must count these outcomes:

$$\underbrace{A \ldots A}\, A$$

$r-1$ A's

k A^c's

2. multiply by $(1-\theta)^k\theta^r$

Negative Binomial, continued

We are counting outcomes with r successes and K failures before before the r^{th} success, that is

Events: $\underbrace{A \ldots A^c \ldots A A^c}_{\substack{r-1 \text{ } A\text{'s} \\ K \text{ } A^c\text{'s}}} \; A$ ← the r^{th} success   prob: $(1-\theta)^K \theta^r$

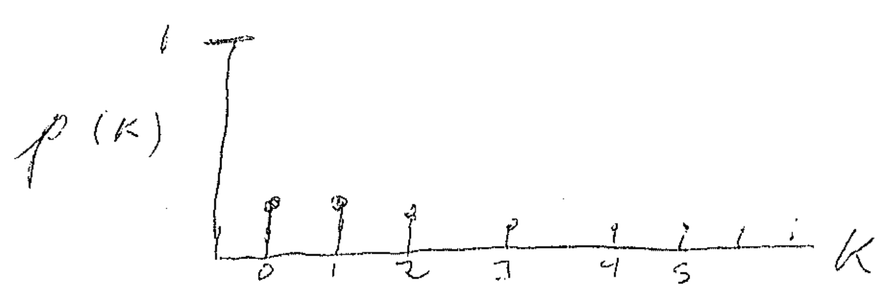
→ there are $(r-1)+K$ positions total in this string of A's and A^c's. We want to choose $r-1$ A's from this string.

This can be done in $\binom{r+K-1}{r-1}$ ways.

Hence: **the** negative binomial distribution

$$nb(K; r, \theta) = Pr(X=k) = \binom{r+K-1}{r-1}(1-\theta)^k \theta^r$$

$$K = 0, 1, 2, \ldots$$



$r = 1$
$\theta = \frac{1}{2}$

Poisson Distribution

Let's take the limit of the Binomial Distribution as $n \to \infty$ while $p \to 0$, but $np = \lambda$

$$p(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

Let $\lambda = np$, so $p = \frac{\lambda}{n}$

$$p(k) = \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!} \frac{n!}{(n-k)!} \frac{1}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}$$

but as $n \to \infty$,

$$\frac{\lambda}{n} \to 0, \quad \frac{n!}{(n-k)!\, n^k} \to 1,$$

and: $\left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}$

$\left(1 - \frac{\lambda}{n}\right)^{-k} \to 1$

$\therefore$

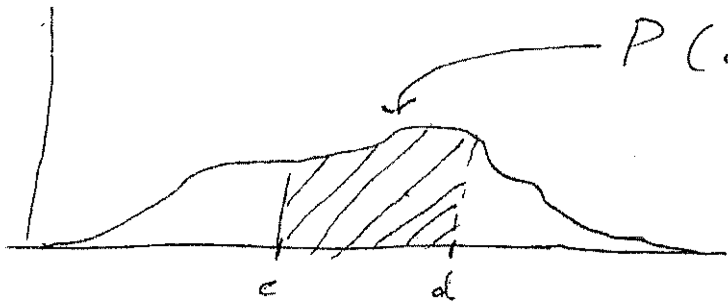
$$p(k) \to \frac{\lambda^k e^{-\lambda}}{k!}$$

Continuous Random Variables

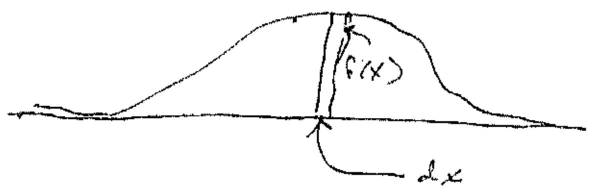
Definition: $f_X(x) = f(x)$ is the <u>probability</u> <u>density</u> of cont. random var X if (a) $f(x) \geq 0$

(b) $\forall c, d \quad c < d$

$$P(c < X \leq d) = \int_c^d f(x)\,dx$$



Neglecting math rigor, we can write

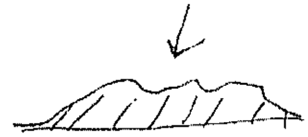


$$f(x)\,dx = P(x < X \leq x + dx)$$

From property (b), the density function f obeys:

$$1 = P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(x)\,dx$$

area = 1



the <u>cumulative</u> <u>distribution</u> <u>function</u>

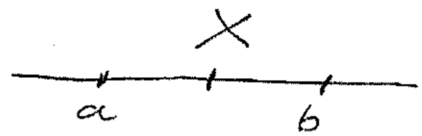
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(s)\,ds$$

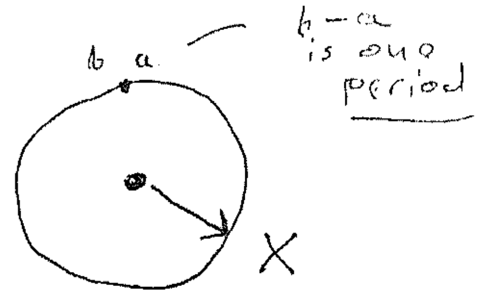
hence

$$\frac{d}{dx} F(x) = f(x)$$

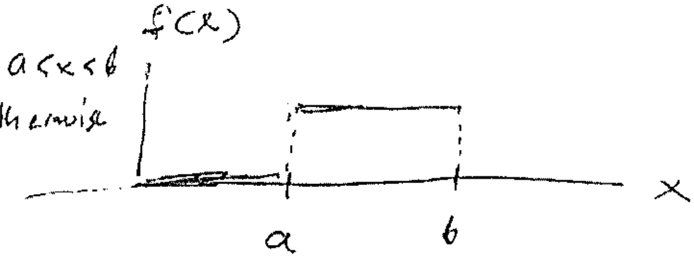
F determines f
f determines F

Example: A spinner (uniform distribution)



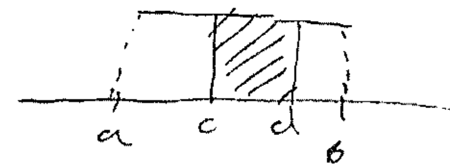


$$f(x) = f(x; a, b) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$



$$P(c < X < d) = \int_c^d f(x)\,dx$$

$$= \frac{d-c}{b-a}$$



Note that

$$P(c < x < d) = P(c \leq x < d)$$
$$= P(c \leq x \leq d) \quad \text{etc etc}$$

so $P(X = c) = 0$, X continuous!

Example: Time before "next event" when events occur with constant probability λ, independently

↳ could be: failure, molecular collision, radioactive decay, etc

Recall that earlier in the lecture, we found the Poisson distribution

$$p(k) = \frac{\lambda^k e^{-\lambda}}{k!}. \quad \text{Then } p(0) = e^{-\lambda}$$

(recall $0! = 1$)

Pick units so that λ is the probability of an event happening between t and $t+1$. Then λ describes a <u>Poisson process</u> with parameter λ. Let an event happen at t_0, and let X be the time to the next event. Then

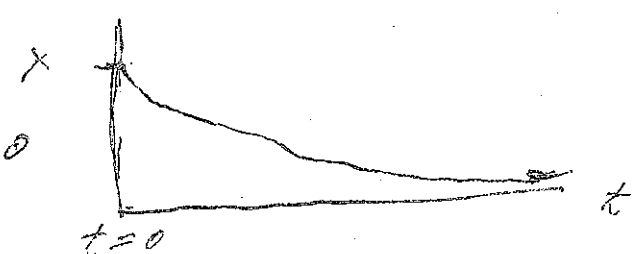
<u>density</u>

$$P(X=t) = f(t) = \begin{cases} \lambda e^{-\lambda t} & (t \geq 0) \\ 0 & (t<0) \end{cases}$$

<u>cumulative</u>

$$P(X<t) = F(t) = \int_{-\infty}^{t} \lambda e^{-\lambda x}\,dx$$

$$= \begin{cases} 1-e^{-\lambda t} & t \geq 0 \\ 0 & t<0 \end{cases}$$



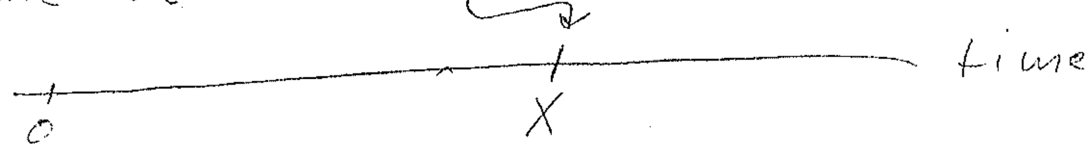
hence time to next event in a Poisson process is governed by the <u>Exponential</u> <u>distribution</u>

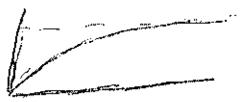
Each event at t_0 is <u>independent</u>, so the system (and the exponential distribution) are said to be "memoryless"

$$P(X>t+s \mid X>s) = P(X>t)$$

What do we mean by "memoryless"?

For concreteness, let X be "time before failure" (say of a lightbulb)



$F(t)$  $F(t) = P(X \le t)$ = Prob. fail by t

$S(t)$  $S(t) = 1 - F(t)$ = Prob. alive at t ("survival function")

$= P(X > t)$

Assume no memory, so

$$P(X > t+s \mid X > s) = P(X > t)$$

$$\frac{P(X > t+s \cap X > s)}{P(X > s)} = P(X > t)$$

$$\frac{P(X > t+s)}{P(X > s)} = P(X > t)$$

$$\frac{S(t+s)}{S(t)} = S(s)$$

$$S(t+s) = S(t)S(s)$$

Obviously, $S(t) = e^{-\lambda t}$ has this property.

Functions of Random Variables

Suppose we have $Y = h(X)$.

Know the distribution of X

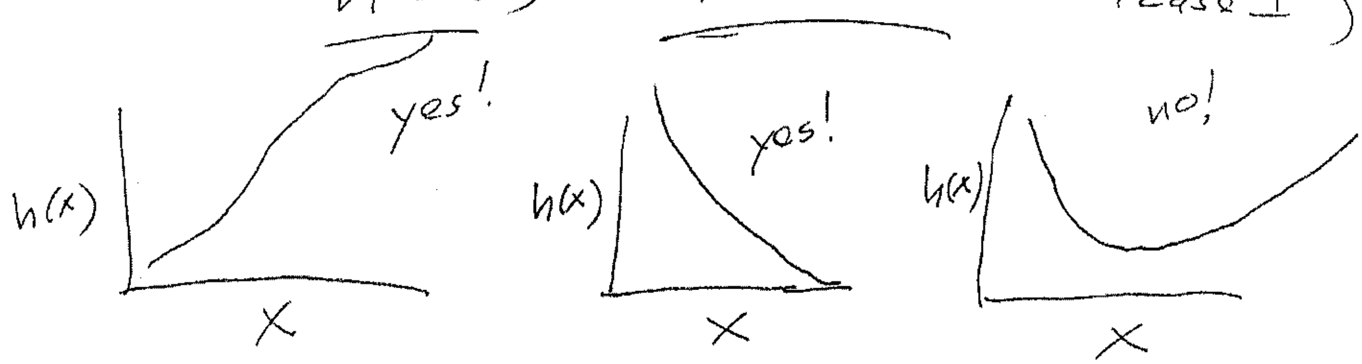
Want the distribution of Y

e.g. I X binomial, want dist of $Y = 2X = h(X)$

II X exponential, want dist of $Y = e^{-\theta x} = h(X)$, etc

$h(X)$ is a coordinate transformation, essentially. Things are most straightforward when:

$h(X)$ monotone (case I)



$Y = \log X$ yes!

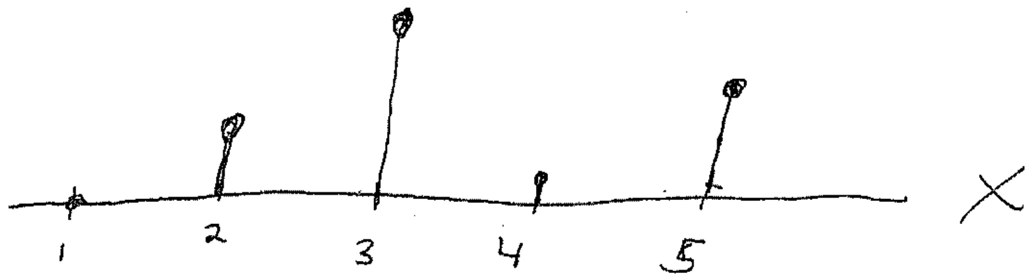
$Y = e^{x}$ yes!

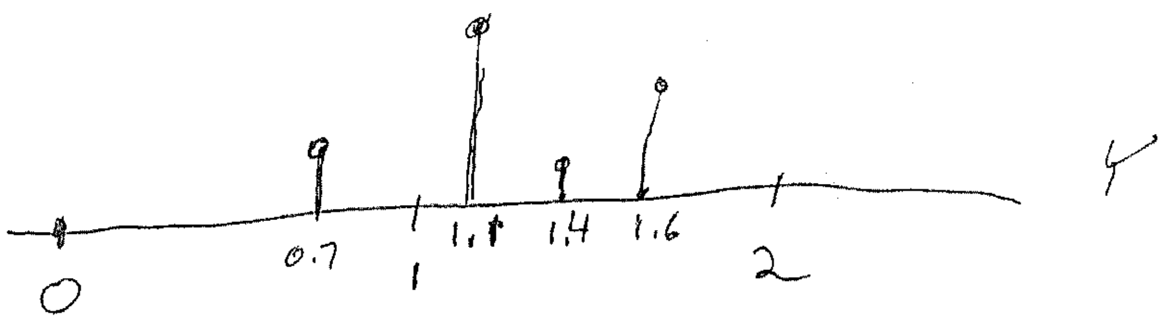
$Y = X^2$ No! $(-\infty < X < \infty)$

[yes if X restricted $(0 < X < \infty)$]

Discrete Case

$$Y = \ln X \quad , \quad X \in \{1, 2, 3, \ldots\}$$



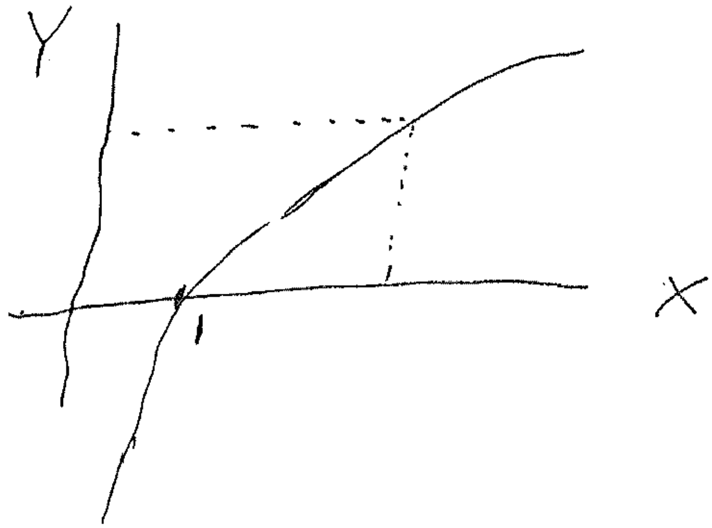


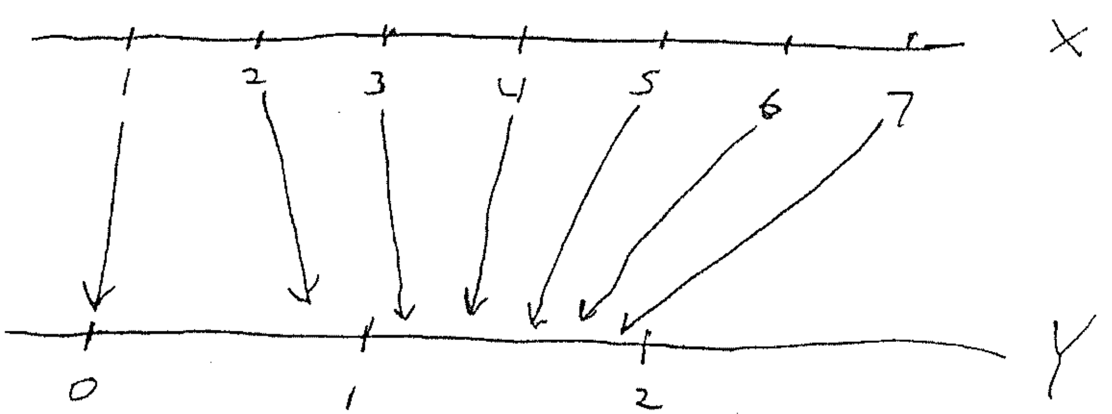
In the discrete case, the only effect of the transformation is to rearrange the spikes. The height of each spike is unchanged.

Basic Idea

Example: $Y = h(X) = \ln X$

so $g(Y) = h^{-1}(Y) = e^{Y} = X$





$\ln(1) = 0$

$\ln(2) = 0.7$

$\ln(3) = 1.1$

$\ln(4) = 1.4$

$\ln(5) = 1.6$

$\ln(6) = 1.8$

$\ln(7) = 1.9$

Functions of Random Variables

Suppose we have $Y = h(X)$.

Know the distribution of X

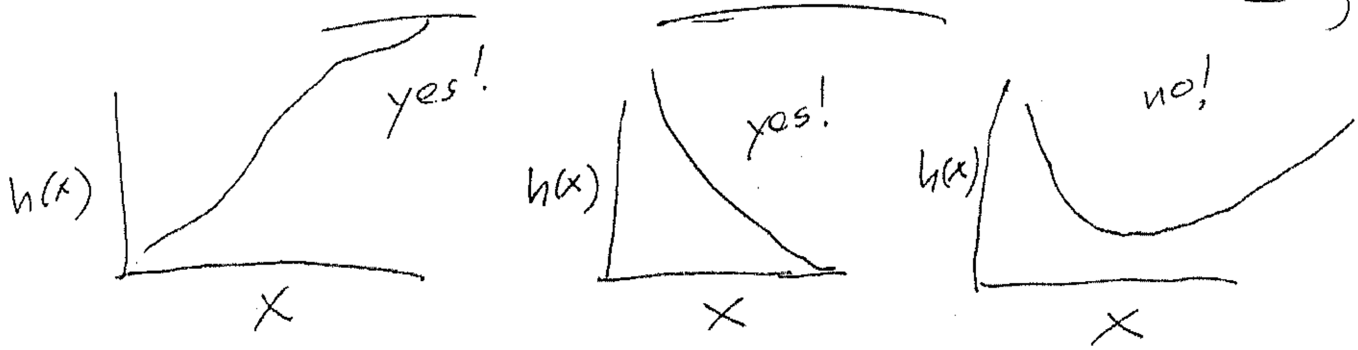
Want the distribution of Y

e.g. I X binomial, want dist of $Y = 2X = h(X)$

II X exponential, want dist of $Y = e^{-\theta x} = h(X)$, etc

$h(X)$ is a coordinate transformation, essentially. Things are most straightforward when:

$h(X)$ monotone (case I)



$Y = \log X$ yes!

$Y = e^x$ yes!

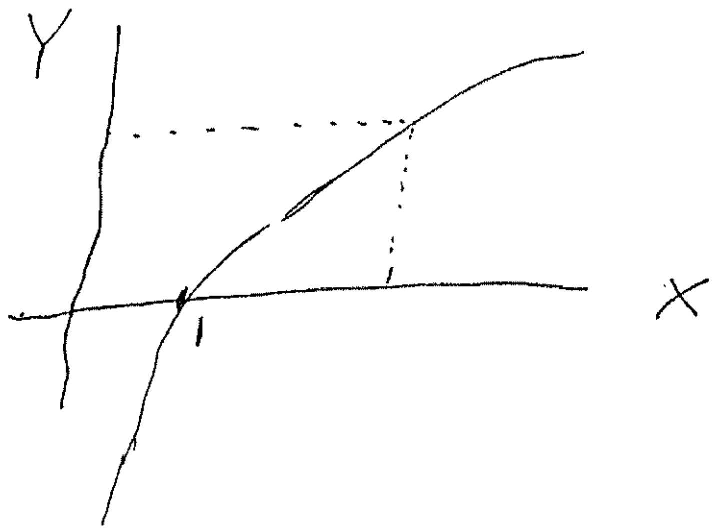
$Y = X^2$ No! $(-\infty < X < \infty)$

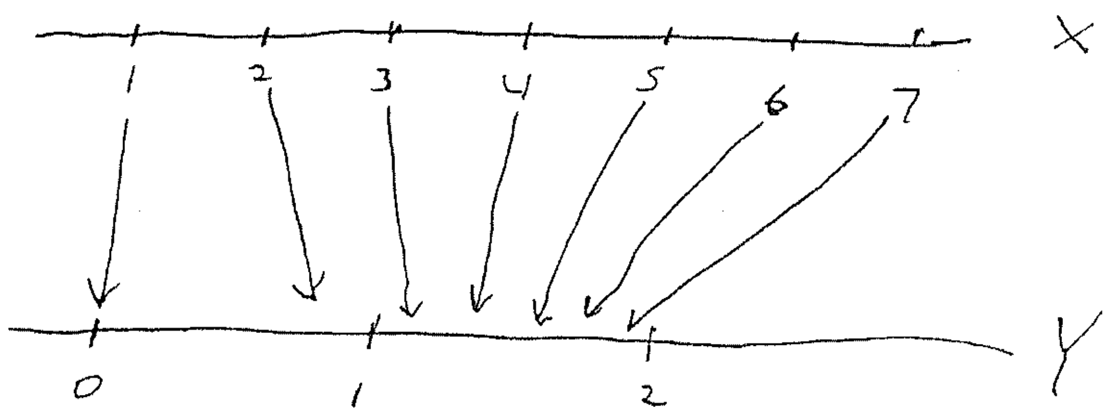
[yes if x restricted $(0 < x < \infty)$]

Basic Idea

Example: $Y = h(X) = \ln X$

$$\text{so } g(Y) = h^{-1}(Y) = e^{Y} = X$$





$\ln(1) = 0$

$\ln(2) = 0.7$

$\ln(3) = 1.1$

$\ln(4) = 1.4$

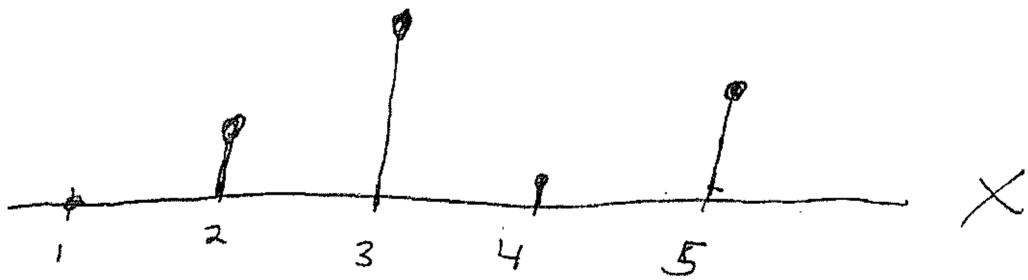
$\ln(5) = 1.6$

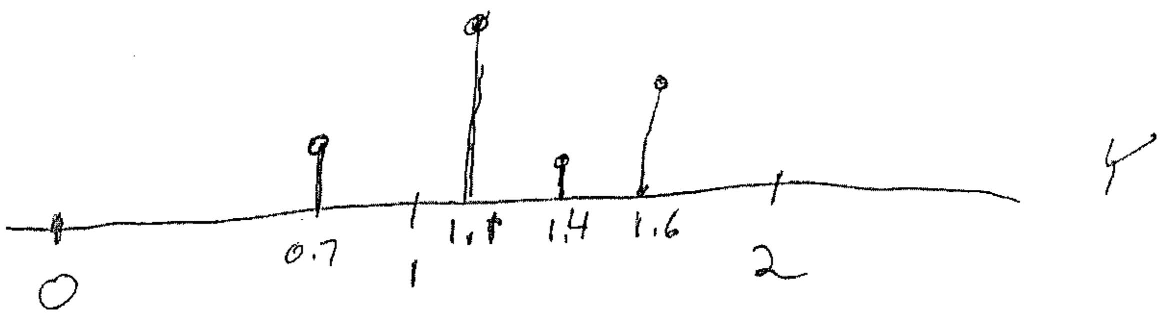
$\ln(6) = 1.8$

$\ln(7) = 1.9$

Discrete Case

$$Y = \ln X \quad , \quad X \in \{1, 2, 3, \ldots\}$$

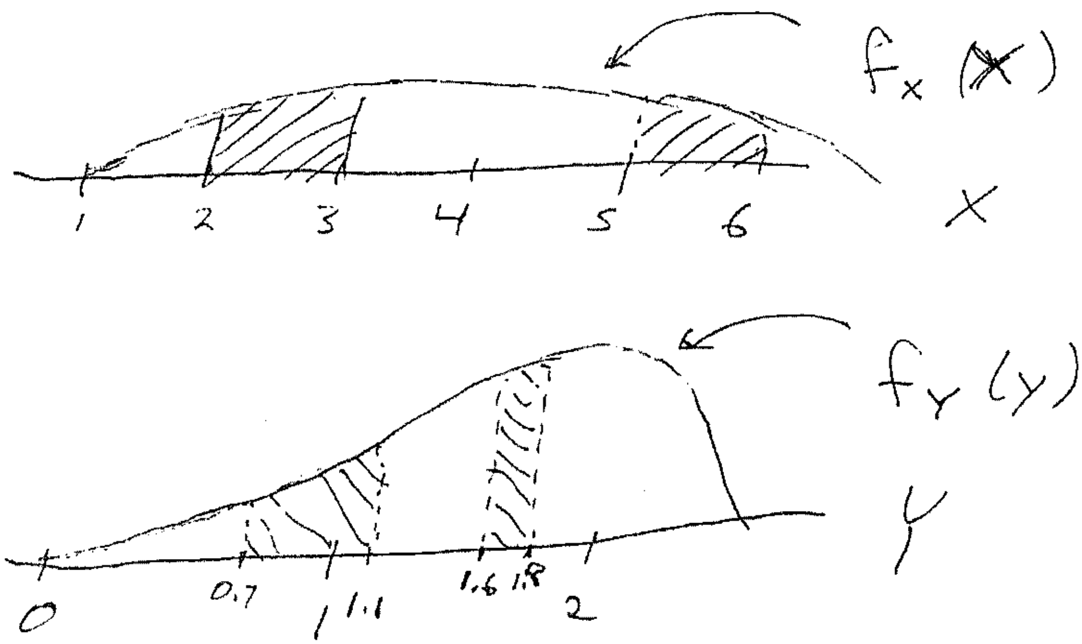




In the discrete case, the only effect of the transformation is to rearrange the spikes. The height of each spike is unchanged.

Continuous Case

$Y = \ln X$ $\quad X$ positive real



The shape of the density changes, but the probability is unchanged. Area represents probability so we must take care to preserve area in the transformation.

If $Y = h(X)$ monotone increasing or decreasing, we just need to solve for x to get

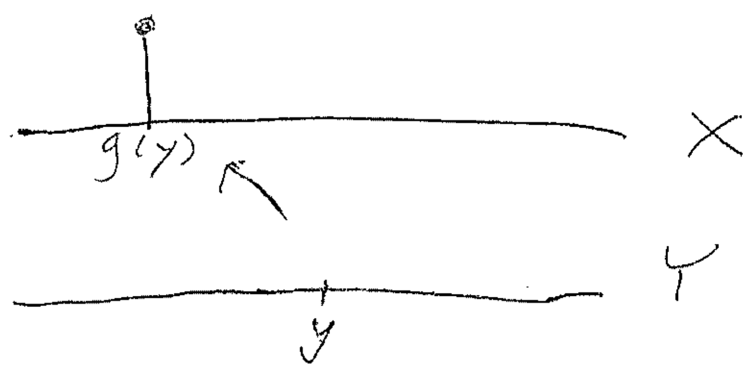
$$X = g(Y)$$

Discrete case: $P_Y(y) = P_X(g(y))$

because:

$$\begin{aligned} P_Y(x) &= P(Y = y) \\ &= P(h(X) = y) \\ &= P(X = g(y)) = P_X(g(y)) \end{aligned}$$

so for each y, to find $P_Y(y)$, find the $x = g(y)$ that led to this g and use its probability,

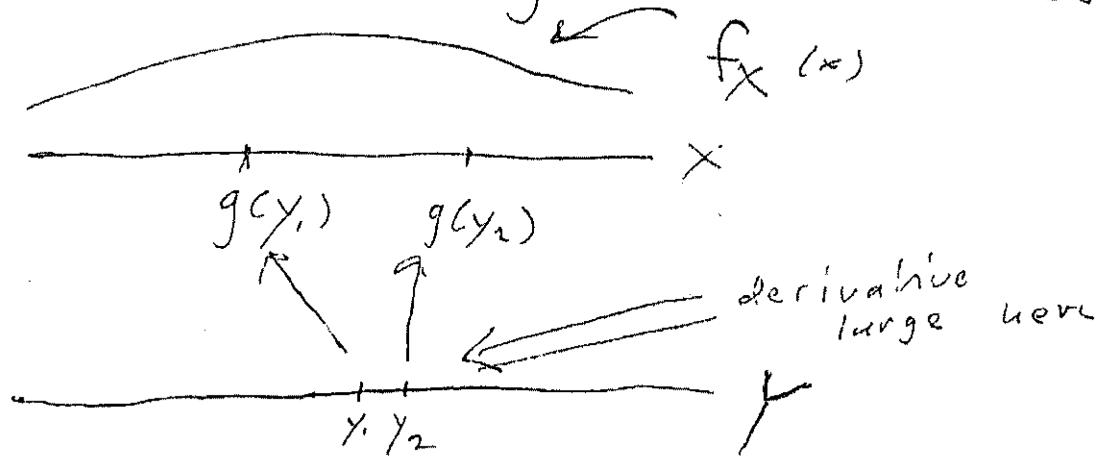


Continuous Case (h monotone)

$$f_Y(y) = f_X(g(y)) \cdot \left| \frac{dg(y)}{dy} \right|$$

rescaling factor to match areas (the "Jacobian")

For each y, to find $f_Y(y)$, "look back" to find the preimage $x = g(y)$ that led to that y, find the density $f_X(g(y))$ at that point; then rescale by $g'(y)$ to take account of how fast g deforms areas.



$$F_Y(a) = P(Y \leq a) = \int_{-\infty}^{a} f_Y(y)\,dy \quad (1)$$

We can also write

$$P(Y \leq a) = P(h(X) \leq a)$$

$$= P(X \leq g(a))$$

$$= \int_{-\infty}^{g(a)} f_X(x)\,dx \quad \left[\begin{array}{l} \text{change:} \\ x = g(y) \\ dx = |g'(y)|\,dy \end{array}\right]$$

$$= \int_{-\infty}^{a} f_X(g(y))\,|g'(y)|\,dy \quad (2)$$

Note the (1) and (2) are equal. Differentiate each side to get

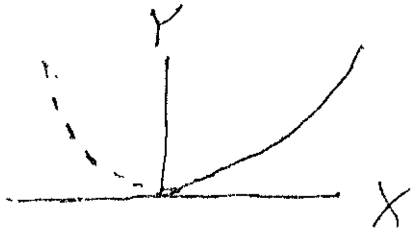
$$f_Y(y) = f_X(g(y))\,|g'(y)|$$

Discrete Example:
X

$$b(x;3,0.5) = \binom{3}{x}(.5)^x(.5)^{3-x} = \binom{3}{x}(.5)^3 = \frac{\binom{3}{x}}{8}$$

$$\left(= \frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}\right) \qquad x \in \{0,1,2,3\}$$

Now let's transform by



$$Y = X^2 \quad \left(\begin{array}{c}\text{monotone} - \\ \underline{\underline{X \geq 0}}\end{array}\right)$$

$$h(x) = x^2 \qquad g(Y) = +\sqrt{Y}$$

$$p_Y(y) = p_X(g(y)) = b(g(y);3,0.5)$$

$$= \begin{cases} \frac{1}{8} & y = 0 \\ \frac{3}{8} & y = 1 \\ \frac{3}{8} & y = 4 \\ \frac{1}{8} & y = 9 \\ 0 & \text{all other } y \end{cases}$$

Continuous Examples:

I. Exponential dist, $\theta > 0$

$$X: f_X(x) = \begin{cases} \theta e^{-\theta x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Suppose X is the time to failure, Y the cost of replacing the part,

$$Y = \frac{1}{1+X}$$

$h(x) = \frac{1}{1+x}$, $1 + x = \frac{1}{y}$, $x = \frac{1}{y} - 1$

so:

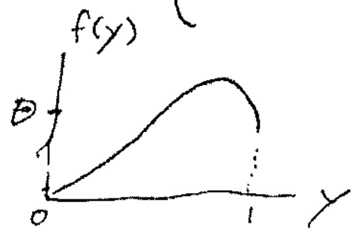
$$g(y) = y^{-1} - 1$$
$$g'(y) = -y^{-2}$$
$$|g'(y)| = y^{-2}$$

Hence $f_Y(y) = f_X\left(\frac{1}{y} - 1\right)\frac{1}{y^2}$

$$= \begin{cases} \theta e^{-\theta\left(\frac{1}{y} - 1\right)}\left(\frac{1}{y^2}\right) & (*) \\ 0 & (**) \end{cases}$$

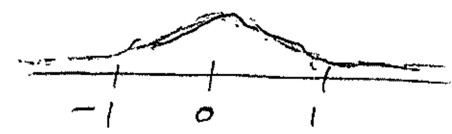
$*$: when $x > 0$ or $0 < y < 1$

$**$: otherwise – $y \leq 0$, $y \geq 1$

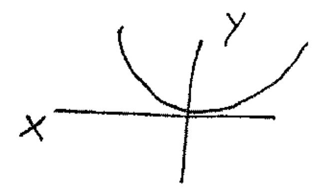


II Standard Normal $N(0,1)$

$$X: f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad -\infty < x < \infty$$



$Y = X^2$ not monotone!! $(-\infty < x < \infty)$



But:

$h(x) = x^2$ has two monotone pieces:

monotone decreasing for $-\infty < x < 0$

monotone increasing for $0 < x < \infty$

Each range has an inverse:

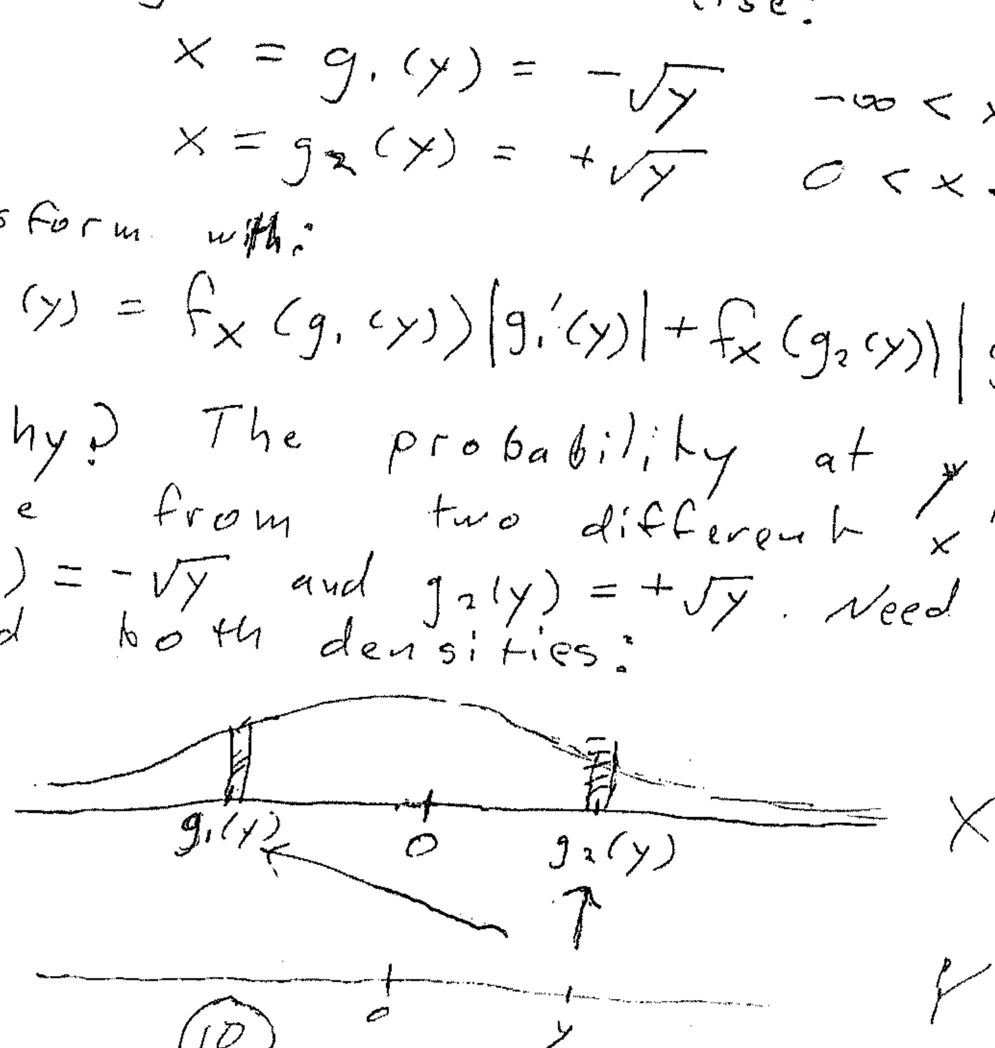
$$x = g_1(y) = -\sqrt{y} \qquad -\infty < x < 0$$

$$x = g_2(y) = +\sqrt{y} \qquad 0 < x < \infty$$

transform with:

$$f_Y(y) = f_X(g_1(y))\,|g_1'(y)| + f_X(g_2(y))\,|g_2'(y)|$$

Why? The probability at y came from two different x's, $g_1(y) = -\sqrt{y}$ and $g_2(y) = +\sqrt{y}$. Need to add both densities:



X standard normal (cont)

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad -\infty < x < \infty$$

$$Y = X^2$$

$$f_Y(y) = f_X(g_1(y))\,|g_1'(y)| + f_X(g_2(y))\,|g_2'(y)|$$

$x = g_1(y) = -\sqrt{y} \quad -\infty < x < 0$; $x = g_2(y) = \sqrt{y} \quad 0 < x < \infty$

$g_1'(y) = \dfrac{-1}{2y^{1/2}}$; $g_2'(y) = \dfrac{1}{2y^{1/2}}$

So

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2} \quad y > 0$$
$$= 0 \quad y \le 0$$

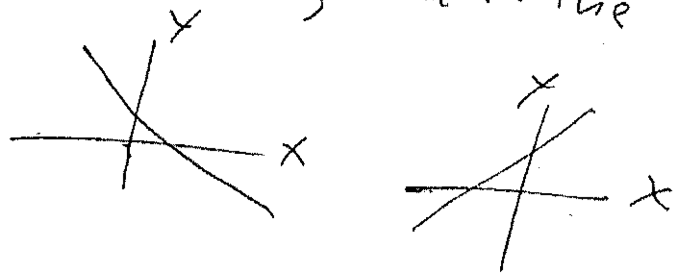
[This is the density function of the χ^2 ("chi-square") distribution with 1 degree of freedom]

Simple VERY IMPORTANT Example

$Y = aX + b$, $a \neq 0$, b constants

("change of scale", "affine")

monotone:



$$X = \frac{Y-b}{a} = g(y), \quad g'(y) = \frac{1}{a}$$

$$|g'(y)| = \frac{1}{|a|}$$

<u>Continuous case</u>:

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right)\frac{1}{|a|}$$

<u>Discrete case</u>:

$$p_Y(y) = p_X\left(\frac{y-b}{a}\right)$$

Example: X standard normal $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$$Y = \sigma X + \mu, \quad \sigma > 0$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

The "general normal" $N(\mu, \sigma^2)$

Example: Suppose X is the time to failure of a light bulb, and we believe X has an exponential(θ) distribution with density

$$f_X(x) = \theta e^{-\theta x} \quad x \geq 0$$
$$= 0 \quad x < 0$$

When the light bulb fails, we replace it with a second one with the same characteristics. The probability the first survives beyond time t, $P(X > t) = e^{-\theta t}$. What is the prob. that

$\boxed{\text{The second bulb survives longer than the first?}}$

That will be $Y = e^{-\theta x} = h(x)$

$$\ln(Y) = -\theta x$$

so $g(Y) = h^{-1}(Y) = \dfrac{-\ln Y}{\theta}$

$h(x) = e^{-\theta x}$ $\quad g(Y) = \dfrac{-\ln Y}{\theta}$

Both monotone decreasing.

$g(Y)$ is only defined for $Y > 0$, but in fact it must be true that $0 < y \leq 1$.

$$g'(y) = \frac{-1}{\theta} \cdot \frac{1}{y}, \text{ and for } y > 0$$

$$|g'(y)| = \frac{1}{\theta y}$$

$$f_Y(y) = f_X(g(y))\,|g'(y)|$$

$f_X(g(y)) = 0$ if $y \leq 0$ or $y \geq 1$, so

$$f_Y(y) = \begin{cases} \theta e^{-\theta\left(\frac{-\ln y}{\theta}\right)} \cdot \dfrac{1}{\theta y} & 0 < \ \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

but $\theta e^{-\theta\left(\frac{-\ln y}{\theta}\right)} = \theta y$, so

$$f_Y(y) = 1 \quad 0 < y < 1$$
$$= 0 \quad \text{otherwise}$$

This is the uniform$(0,1)$ distribution

The Probability Integral Transform

More generally, if X is a continuous random variable, and

$$Y = F(x), \quad \text{then}$$

$\boxed{\text{The cdf of } x}$

The cdf of Y is

$$P(Y \le y) = P(F(x) \le y)$$
$$= P(F^{-1}(F(x)) \le F^{-1}(y))$$
$$= P(x \le F^{-1}(y))$$
$$= y \quad (0 \le y \le 1)$$

$F(y) = y$, the cdf of the Uniform distribution

because

$$f_y(y) = \frac{dF}{dy} = 1 \quad 0 \le y < 1$$

See Rice, pp 62-63. This result is very useful for generating random deviates.

STAT 244
Lecture 4
10/6/16

Describing Probability Distributions of Random Vars

Full description: $p_X(x)$ or $f(x)$

Partial, much shorter, description:

Measures of center and dispersion:

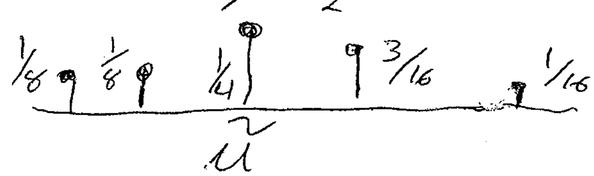
Center: Expectation ("expected value" "mean")

Median (implies "middle")

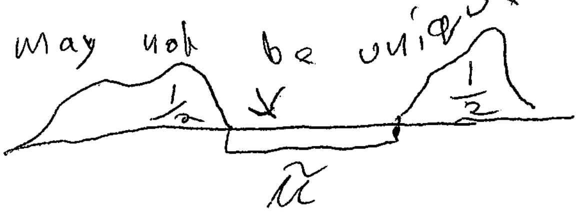
First, consider median:

want a $\tilde{\mu}$ such that $F(\tilde{\mu}) = \frac{1}{2}$,

so $P(X \geq \tilde{\mu}) > \frac{1}{2}$, $P(X \leq \tilde{\mu}) \leq \frac{1}{2}$



may not be unique

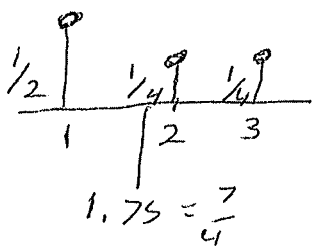


Expectation of X

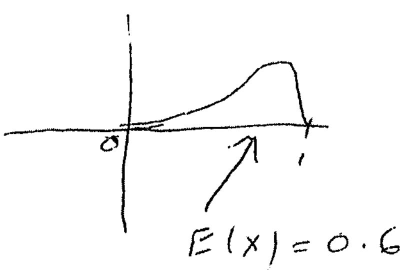
"Center of Gravity", 'Fair Price'

Def $E(X) = \begin{cases} \sum_{\text{all } x} x\, p_X(x) & \text{Discrete} \\ \int_{-\infty}^{\infty} x\, f_X(x)\, dx & \text{continuous} \end{cases}$

Simple Examples:



$$E(X) = 1 \cdot \tfrac{1}{2} + 2 \cdot \tfrac{1}{4} + 3 \cdot \tfrac{1}{4} = \tfrac{7}{4}$$



$$f_X(x) = \begin{cases} 12x^2(1-x) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx = \int_0^1 x\; 12x^2(1-x)\, dx$$

$$= 12 \int_0^1 (x^3 - x^4)\, dx = 12 \left[\frac{x^4}{4} - \frac{x^5}{5} \right|_0^1$$

$$= \frac{12}{20} = 0.6$$

$$E(X) = \begin{cases} \sum_{\text{all } x} x\, p_x(x) & \underline{\text{discrete}} \\ \int_{-\infty}^{\infty} x\, f_x(x)\,dx & \underline{\text{continuous}} \end{cases}$$

So... suppose we have a function of a random variable $h(X)$.

What is $E(X^2)$? $E(e^X)$? $E(\ln X)$?

Is $E(h(x)) = h(E(X))$? (~~no~~, usually <u>NOT</u>)

$Y = h(X)$. 2 ways to find $E(Y)$.

① Find $f_Y(y)$, $E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)\,dy$

② $E(Y) = \int_{-\infty}^{\infty} h(x)\, f_X(x)\,dx$ (discrete: $E(Y) = \sum h(x)\, p_x(x)$)

Usually ② is easier.

<u>Pf</u> $E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)\,dy = \int_{-\infty} y\, f_x(g(y))\,g'(y)\,dy$

but $x = g(y)$ $y = h(x)$ $dx = g'(y)dy$

so

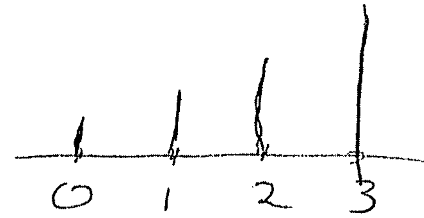
$$E(Y) = \int_{-\infty}^{\infty} h(x)\, f_x(x)\,dx$$

③

Examples

①

X	0	1	2	3
$p_X(x)$	.1	.2	.3	.4

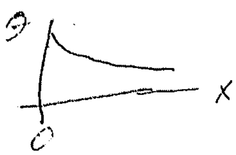


$$E(X) = 0(0.1) + 1(0.2) + 2(0.3) + 3(0.4)$$
$$= \boxed{2}$$

$$E(X^2) = 0^2(0.1) + 1^2(0.2) + 2^2(0.3) + 3^2(0.4)$$
$$= \boxed{5}$$

NOTE: $E(X^2) \neq (E(X))^2$

② X exponential, $f_X(x) = \theta e^{-\theta x}$, $x > 0$



$$E(X) = \int_0^\infty x\,\theta e^{-\theta x}\,dx = \frac{1}{\theta}\int_0^\infty u e^{-u}\,du$$ (ch. vars: $u = \theta x$, $\frac{1}{\theta}du = dx$)

$$\int_0^\infty u e^{-u} = -ue^{-u}\Big|_0^\infty + \int_0^\infty e^{-u}\,du$$ (integration by parts)

$$= 0 + -e^{-u}\Big|_0^\infty = 1$$

So $E(X) = \left(\frac{1}{\theta}\right)(1) = \frac{1}{\theta}$

$$E(X^2) = \int_0^\infty x^2\theta\, e^{-\theta x}\,dx = \frac{1}{\theta^2}\int_0^\infty u^2 e^{-u}\,du$$

$$\int_0^\infty u^2 e^{-u}\,du = -u^2e^{-u}\Big|_0^\infty + \int_0^\infty 2u e^{-u}\,du$$
$$= 2\int_0^\infty u e^{-u}\,du = 2$$ (see above)

So $E(X^2) = \frac{2}{\theta^2} \neq E(X)$

General Examples

① **Linear Transformations** $h(x) = aX + b$

Then $E(aX+b) = aE(X) + b$

(a case where $E(h(X)) = h(E(X))$)

Pf $E(aX+b) = \int_{-\infty}^{\infty} (ax+b) f_X(x)\,dx$

$$= \int_{-\infty}^{\infty} (a x f_X(x) + b f_X(x))\,dx = a \underbrace{\int_{-\infty}^{\infty} x f_X(x)\,dx}_{E(X)} + b \underbrace{\int_{-\infty}^{\infty} f_X(x)\,dx}_{1}$$

$\boxed{= aE(X) + b}$ ∴

Note that if $a = 0$, get $E(b) = b$,
True for all constants including $E(X)$

$E(E(X)) = E(X)$

② **Variances** $h(X) = (X - \mu_X)^2$

Def: $Var(X) = E[(X-\mu_X)^2]$ is variance of X (or of prob dist of X)

Notation: $Var(X) = \sigma_X^2$ (or just σ^2 when no confusion)

$\sqrt{Var(X)} = \sigma_X$ the standard deviation of X

var and s.d. measure dispersion, spread.

For Theoretical Calculation:

$$\boxed{Var(X) = E(X^2) - \mu_x^2}$$

Pf: $Var(X) = E\left[(x - \mu_x)^2\right]$

$$= \int_{-\infty}^{\infty} (x^2 - 2x\mu_x + \mu_x^2) f_X(x)\,dx$$

$$= \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - 2\mu_x \int_{-\infty}^{\infty} x f_X(x)\,dx + \mu_x^2 \int_{-\infty}^{\infty} f_X(x)\,dx$$

$$= E(x^2) - 2\mu_x \cdot \mu_x + \mu_x^2 \cdot 1$$

$$= E(X^2) - \mu_x^2$$

Examples: Exponential dist

$$E(x^2) = \frac{2}{\theta^2}, \quad E(x) = \frac{1}{\theta}$$

$$Var(x) = \frac{2}{\theta^2} - \frac{1}{\theta^2} = \boxed{\frac{1}{\theta^2}}$$

Linear Transformation:

$$E(aX + b) = a\mu_x + b$$

$$Var(aX + b) = a^2 Var(X)$$

(so $\mu_{ax+b} = a\mu_x + b$; $\sigma^2_{ax+b} = a^2 \sigma_x^2$)

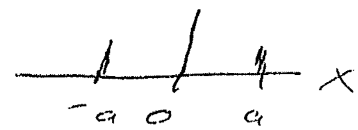
Special Case: $w = \frac{x - \mu_x}{\sigma_x}$ // Standard Form

$$E(W) = \frac{E(X) - \mu_x}{\sigma_x} = 0 \; ; \; Var(W) = \sigma_w^2 = 1$$

Interpreting Variance

example:

x	$-a$	0	a
$p_X(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$



$E(X) = 0$ (by symmetry, or

$-a(\frac{1}{4}) + 0(\frac{1}{2}) + a(\frac{1}{4}) = 0$)

so $\mathrm{Var}(X) = E(X^2)$

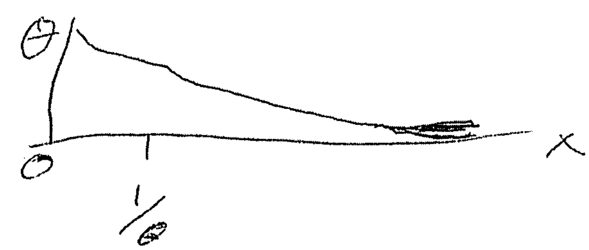
$= (-a)^2(\frac{1}{4}) + 0^2(\frac{1}{2}) + a^2(\frac{1}{4})$

$= \frac{a^2}{2}$

large $a \Leftrightarrow$ large $\mathrm{Var}(X) \Leftrightarrow$ large spread

note: Squared units!

Exponential:



$E(X) = \frac{1}{\theta}$

$\mathrm{Var}(X) = \frac{1}{\theta^2}$, $\sigma_X = \frac{1}{\theta}$

Multivariate, or Joint Distributions

Distributions of 2, 3, 4, or more random vars.

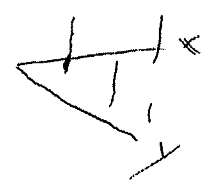
Discrete Bivariate Case:

X, Y are 2 rand. vars. defined on same sample space.

Bivariate prob. function:

$$p(x,y) = Pr(X = x \text{ and } Y = y)$$

$$\sum_{\text{all } x} \sum_{\text{all } y} p(x,y) = 1$$



Ex; 1 Toss 2 fair coins 3 times each

$X = \#H\text{'s}$ coin 1

$Y = \#T\text{'s}$ coin 2

$Z = \#T\text{'s}$ coin 1

dist of (X, Y)

$x \backslash y$	0	1	2	3
0	1/64	3/64	3/64	1/64
1	3/64	9/64	9/64	3/64
2	3/64	9/64	9/64	3/64
3	1/64	3/64	3/64	3/64

dist of (X, Z)

$x \backslash z$	0	1	2	3
0	0	0	0	1/8
1	0	0	3/8	0
2	0	3/8	0	0
3	1/8	0	0	0

Can compute univariate prob. dist. from bivariate distributions by addition:

$$p_X(x) = \sum_{\text{all } y} p(x, y)$$

$$p_Y(y) = \sum_{\text{all } x} p(x, y)$$

These are called the marginal prob functions of X and Y respectively.

Idea: The event

$$\{X = x\} = \{X = x \text{ and } Y = 1\} \cup \{X = x \text{ and } Y = 2\} \ldots$$

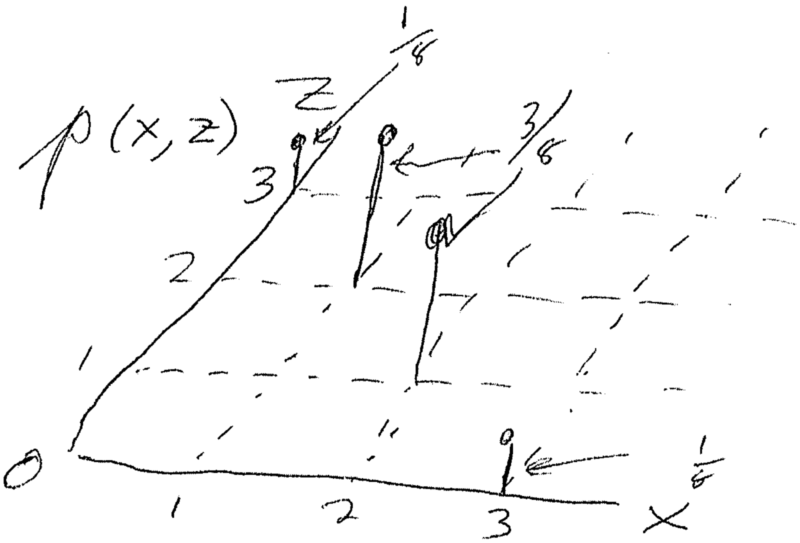
$$\ldots \cup \{X = x \text{ and } Y = 57\} \ldots$$

The events on the right side are mutually exclusive, so we can add probabilities

3 coin example, part 2:

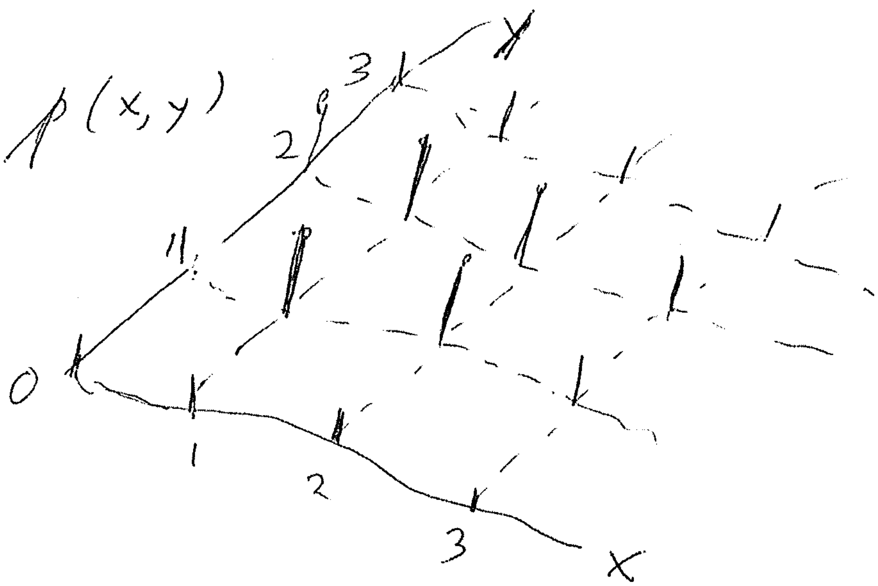
X \ Y	0	1	2	3	
0	1/64	3/64	3/64	1/64	1/8
1	3/64	9/64	9/64	3/64	3/8
2	3/64	9/64	9/64	3/64	3/8
3	1/64	3/64	3/64	1/64	1/8
	1/8	3/8	3/8	1/8	1

X \ Z	0	1	2	3	
0	0	0	0	1/8	1/8
1	0	0	3/8	0	3/8
2	0	3/8	0	0	3/8
3	1/8	0	0	0	1/8
	1/8	3/8	3/8	1/8	1

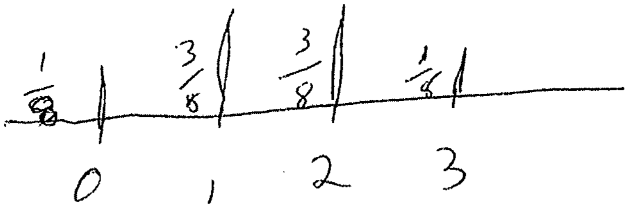


$Z = 3 - X$

not indep!



X, Y indep.



marginal dist is like a "side view" of joint dist.

IMPORTANT

Conditional Prob. Functions

$$p(y/x) = Pr(\text{"}Y=y\text{"} \,|\, \text{"}X=x\text{"}) \quad \text{"given"}$$

$$= \frac{Pr(\text{"}X=x\text{" AND "}Y=y\text{"})}{Pr(\text{"}X=x\text{"})}$$

$$p(y/x) = \frac{p(x,y)}{p_x(x)}, \quad \sum_{\text{all } y} p(y/x) = 1$$

If $p(y/x) = p_Y(y)$ for all x, y

(ie $p(y/x) = \frac{p(x,y)}{p_x(x)}$, so $p(x,y) = p_x(x)\,p_Y(y)$)

we say the random variables X and Y are independent

In the previous example
X and Y are indep
X and Z are dependent

This is true even though they have the same marginal distribution.

Discrete Example:

$X = \#H$'s
$Z = \#T$'s
} Same coin

$X + Z = 3$

$X \backslash Z$	0	1	2	3
0	0	0	0	$\frac{1}{8}$
1	0	0	$\frac{3}{8}$	0
2	0	$3/8$	0	0
3	$\frac{1}{8}$	0	0	0
	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

"Marginal" $p_Z(z)$:

z	0	1	2	3
p	$1/8$	$3/8$	$3/8$	$1/8$

"Conditional" $p(x|z)$, say for $Z=2$:

$$p(x|Z=2) = \frac{p(x,2)}{p_Z(2)}$$

x	0	1	2	3
$p(x\|z=2)$	0	1	0	0

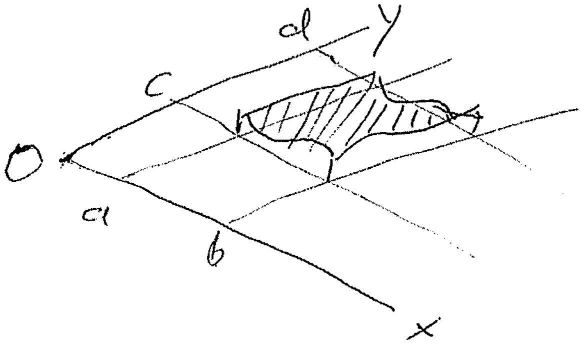
Continuous bivariate case

Bivariate prob. density:

① $f(x,y) \geq 0$

② $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$ (a double integral)

$$\Pr(a < X < b \text{ and } c < Y < d) = \underbrace{\int_c^d \int_a^b f(x,y)\,dx\,dy}_{\text{volume between the x-y plane and the surface } f(x,y)}$$



Marginal prob. densities

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

Conditional prob densities

$$f(y/x) = \frac{f(x,y)}{f_X(x)}$$

IMPORTANT

If $f(y/x) = f_Y(y)$ for all x, y, X and Y are independent and $f(x,y) = f_X(x)\,f_Y(y)$

Interpreting Conditional Densities

Intuitively, think of

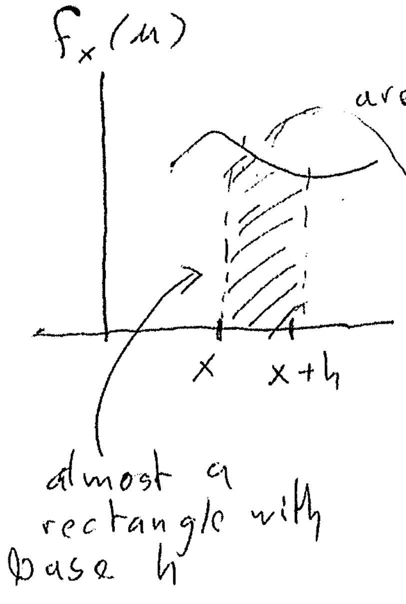
$$f(y/x)\,dy = \Pr(y \le Y \le y + dy \mid X = x)$$

even though the right hand side is not defined (since $\Pr(X = x) = 0$)

Under regularity conditions, one can get (for h small)

$$\Pr(a < Y \le b \mid X = x) \simeq \Pr(a < Y \le b \mid x \le X \le x + h)$$

$$= \frac{\Pr(a < Y \le b \text{ AND } x \le X \le x + h)}{\Pr(x \le X \le x + h)}$$



$$= \frac{\int_a^b \int_x^{x+h} f(u, y)\,du\,dy}{\int_x^{x+h} f_X(u)\,du}$$

$$\simeq \frac{\int_a^b f(x, y)\,h\,dy}{f_X(x)\,h}$$

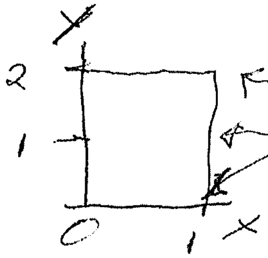
$$= \int_a^b \left[\frac{f(x, y)}{f_X(x)}\right] dy$$

conditional probability by area

Example X, Y random variables

joint (bivariate) density

$$f(x,y) = \begin{cases} y\left(\frac{1}{2} - x\right) + x & \text{for } 0 < x < 1 \\ & \phantom{\text{for }} 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}$$

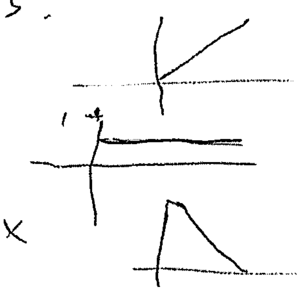


cross-sections:

$f(x,0) = x$

$f(x,1) = \frac{1}{2}$

$f(x,2) = 1 - x$



$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \iint \left[y\left(\tfrac{1}{2} - x\right) + x\right] dx\,dy$$

$$= \int_0^2 \left[\int_0^1 \left[y\left(\tfrac{1}{2} - x\right) + x\right] dx\right] dy$$

$$= \int_0^2 \left[\int_0^1 y\left(\tfrac{1}{2} - x\right) dx + \int_0^1 x\,dx\right] dy$$

$$= \int_0^2 \left[y \int_0^1 \left(\tfrac{1}{2} - x\right) dx + \int_0^1 x\,dx\right] dy$$

$$= \int_0^2 \left[y \cdot 0 + \frac{1}{2}\right] dy$$

$$= \int_0^2 \frac{1}{2}\,dy = \frac{1}{2} \cdot 2 = 1$$

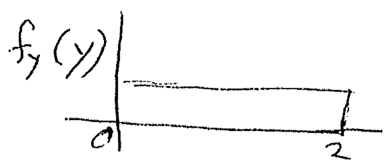
In the process, we found

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^1 \left[y\left(\tfrac{1}{2}-x\right)+x\right]dx = \tfrac{1}{2}$$

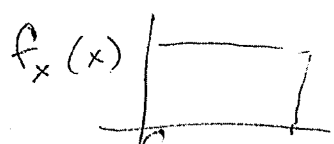
$\rightarrow$ This is $f_Y(y) = \begin{cases} \tfrac{1}{2} & 0 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}$

ie, the marginal of Y is <u>uniform</u>.

Similarly $f_X(x) = \int_0^2 \left(y\left(\tfrac{1}{2}-x\right)+x\right)dy$



$$= \frac{y^2}{2}\Big|_0^2 \left(\tfrac{1}{2}-x\right) + 2x = 1 + 2x - 2x$$

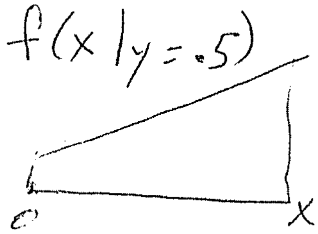


$$= \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

also uniform!

but X, Y not indep:

$$f(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{y\left(\tfrac{1}{2}-x\right)+x}{\left(\tfrac{1}{2}\right)}$$



$$= \begin{cases} y(1-2x)+2x & 0 \le x \le 1 \\ 0 & \end{cases}$$

depends on y!

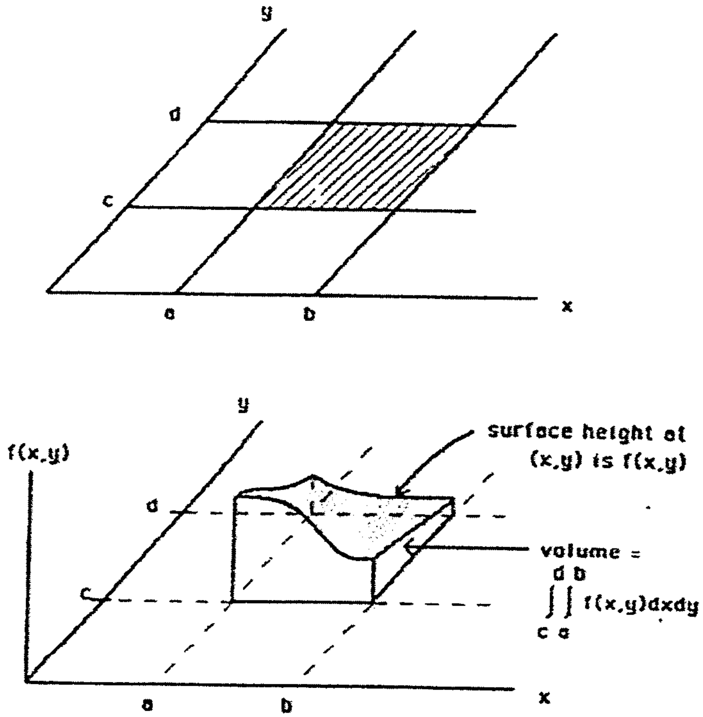


Figure 3.3. A bivariate density.

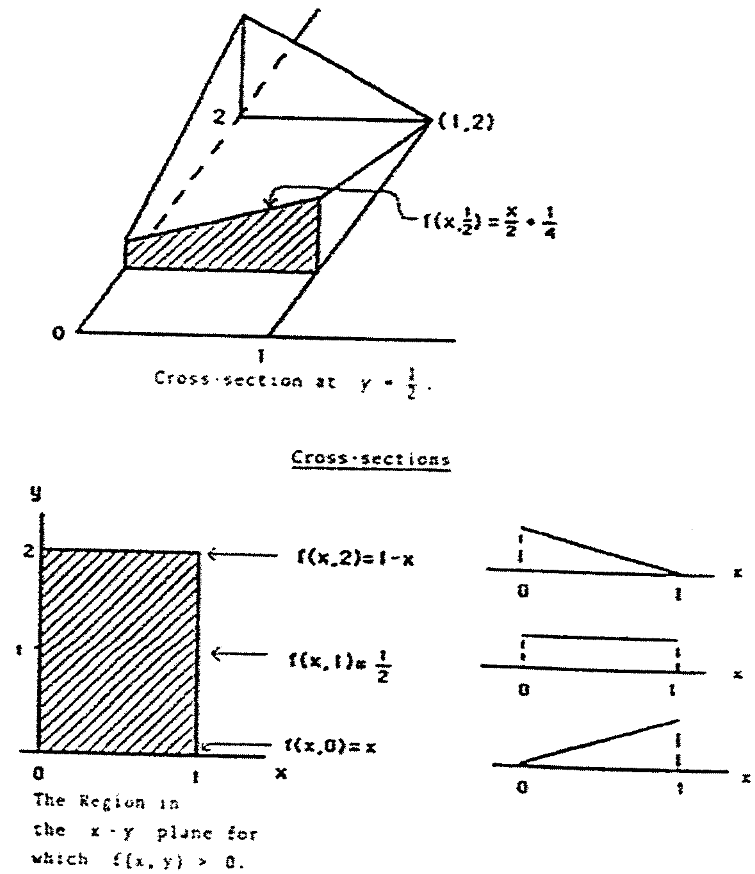


Figure 3.4. A bivariate density and its cross-sections.

STAT 24400
Lecture 5
October 11, 2016

We've talked about expectations, and we've talked about joint distributions. What are expectations of joint distributions $E[h(X,Y)]$?

As with a single variable: we could ① Let $Z = h(X,Y)$, a random var, and then find $f_Z(z)$, find $\int_{-\infty}^{\infty} z f_Z(z)\,dz$

(this is hard)

or

② Find $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) f(x,y)\,dx\,dy$

In general, we sidestep the issue and define

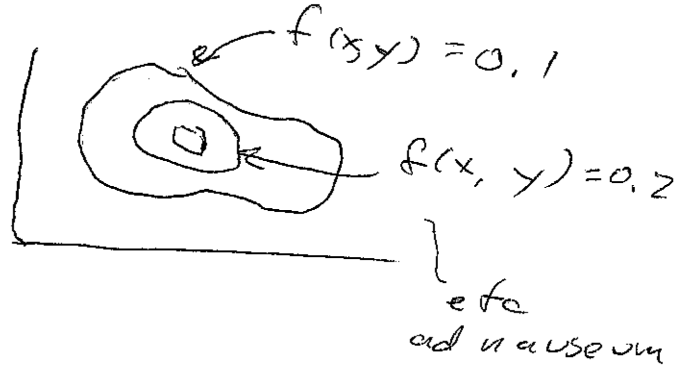
$$E(h(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) f(x,y)\,dx\,dy$$

$$E(h(X_1, X_2, \dots, X_n)) = \underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{n\text{-fold}} h(x_1,\dots,x_n) f(x_1,\dots x_n)\,dx_1\dots dx_n$$

Describing Multivariate Distributions

One Way: $f(x, y)$

Hard to graph, and for more than 2D very hard to graph.



Another Way: single number summaries, covariance, correlation

---

Recall $E(h(X, Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) f(x,y)\, dx\, dy$

for $h(x,y) = x + y$

$$E(X+Y) = \iint (x+y) f(x,y)\, dx\, dy$$

$$= \iint x f(x,y)\, dx\, dy + \iint y f(x,y)\, dx\, dy$$

$$= \int x \underbrace{\left[\int f(x,y)\, dy\right]}_{f_X(x)} dx + \int y \underbrace{\left[\int f(x,y)\, dx\right]}_{f_Y(y)} dy$$

$$= \int x f_X(x)\, dx + \int y f_Y(y)\, dy$$

$$= E(X) + E(Y)$$

More Generally: For any random variables $X_1, X_2, \ldots, X_n$

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

hence: $$E\left(\sum_{i=1}^{n} h_i(X_i)\right) = \sum_{i=1}^{n} E(h_i(X_i))$$

In particular, expectations of linear functions are linear functions of <u>marginal</u> <u>expectations</u>:

$$E(aX + bY) = aE(X) + bE(Y)$$

marginal, univariate

Since marginal distributions do not in general determine a bivariate distribution, we <u>cannot</u> describe bivariate dists (only) with <u>expectations</u> of linear funcs

need more:

Covariance of X and Y

$$\operatorname{cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

$$[\,] = XY - \mu_x Y - \mu_y X + \mu_x \mu_y$$

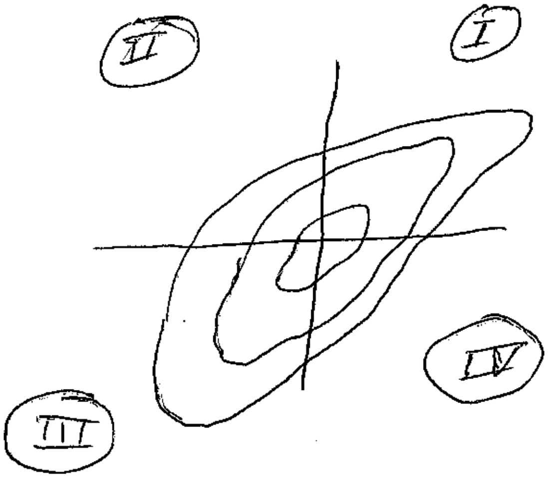
$$E[\,] = E[XY] - \mu_x \underbrace{E(Y)}_{\mu_y} - \mu_y \underbrace{E(X)}_{\mu_x} + \mu_x \mu_y$$

$$= E[XY] - \mu_x \mu_y$$

$$\boxed{\operatorname{cov}(X, Y) = E[XY] - \mu_x \mu_y}$$

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)\,dx\,dy$$

note that if $E(X)=E(Y)=0$



$$Cov(X,Y)=E(XY)$$

$E(XY)>0$ if more prob. in $(I + III)$ than $(II + IV)$

note further that $Cov(X,Y)=E[(X-\mu_x)(Y-\mu_y)]$ implies that

$$Cov(X,X)=Var(X)$$

Other properties:

$$Cov(a+bX,\ c+dY)=bd\,Cov(X,Y)$$

If X,Y indep. $Cov(X,Y)=0$

__But__ can have $Cov(X,Y)=0$ with X, Y dependent

__Correlation__: "scale free" covariance

$$corr(X,Y)=\rho_{XY}=Cov\left(\frac{X-\mu_x}{\sigma_x},\frac{Y-\mu_y}{\sigma_y}\right)=\frac{Cov(X,Y)}{\sigma_x\sigma_y}$$

__also called__ "Pearson's Correlation Coefficient"

Example:
$$f(x,y)=\begin{cases}\frac{4}{5}(xy+1) & 0<x<1 \\ & 0<y<1 \\ 0 & \text{otherwise}\end{cases}$$

Is it really a distribution? Let's check!

$$\int_0^1\int_0^1 f(x,y)\,dx\,dy = \int_0^1 \underbrace{\left(\int_0^1 \frac{4}{5}(xy+1)\,dx\right)}_{f_Y(y)} dy$$

$$= \int_0^1 \underbrace{\frac{4}{5}\left(\frac{1}{2}y+1\right)}_{f_Y(y)} dy = \frac{4}{5}\left(\frac{y^2}{4}+y\right)\Big|_0^1 = 1$$

$$E(Y)=\int_0^1 y\, f_Y(y)\,dy = \frac{4}{5}\int_0^1\left(\frac{1}{2}y^2+y\right)dy = \frac{8}{15}$$

similarly $E(X) = 8/15$

$$E(XY)=\int_0^1\int_0^1 xy\, f(x,y)\,dx\,dy = \int_0^1\left[\int_0^1 \frac{4}{5}(x^2y^2+xy)\,dx\right]dy$$

$$= \int_0^1 \frac{4}{5}\left[\frac{y^2}{3}+\frac{y}{2}\right]dy = \frac{4}{5}\left[\frac{y^3}{9}+\frac{y^2}{4}\right]_0^1 = \frac{13}{45}$$

$$\Rightarrow \text{Cov}(X,Y) = \frac{13}{45} - \left(\frac{8}{15}\right)\left(\frac{8}{15}\right) = \frac{1}{225}$$

Can find ρ_{xy}. But what does the magnitude of covariance mean?

To interpret magnitude of Cov, consider $X+Y$.

$$E(X+Y) = E(X) + E(Y) = \mu_X + \mu_Y$$

$$\begin{aligned}
Var(X+Y) &= E\left[(X+Y) - E(X+Y)\right]^2 \quad \text{(def)} \\
&= E\left[(X+Y) - (\mu_X + \mu_Y)\right]^2 \\
&= E\left[(X-\mu_X) + (Y-\mu_Y)\right]^2 \\
&= E\left[(X-\mu_X)^2 + (Y-\mu_Y)^2 + 2(X-\mu_X)(Y-\mu_Y)\right] \\
&= Var(X) + Var(Y) + 2\,Cov(X,Y)
\end{aligned}$$

General Interp. of Covariance: It is a <u>correction factor</u> for finding variances of sums

$$\Rightarrow \boxed{Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)}$$

So:

$$Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$$

if $Cov(X,Y) = 0$ ("X, Y uncorrelated")

then $Var(X+Y) = Var(X) + Var(Y)$

More generally:

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + \sum_{i}\sum_{j}_{i \neq j} Cov(X_i, X_j)$$

using

$$Var(aX) = a^2 Var(X)$$
$$Cov(aX, bY) = ab\,Cov(X,Y)$$

we get

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i) + \sum\sum_{i \neq j} a_i a_j Cov(X_i, X_j)$$

If each pair X_i and X_j $i \neq j$ are uncorrelated

$$Var\left(\sum a_i X_i\right) = \sum a_i^2 Var(X_i)$$

Let's continue with the case

$Cor(X_i, X_j) = 0$ iff $i \neq j$, so

$$Var(\sum a_i X_i) = \sum a_i^2 Var(X_i)$$

introduce a new random var

$$\bar{X} = \sum \frac{1}{n} X_i \qquad (\text{so } a_i = \tfrac{1}{n})$$

Then $Var(\bar{X}) = \frac{1}{n^2} \sum Var(X_i)$

now suppose the X_i are identically distributed, so

$Var(X_i) = \sigma^2$ for all i

then

$$\underline{\underline{Var(\bar{X}) = \frac{\sigma^2}{n}}}$$

Note that $E(\bar{X}) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \mu_x$

Recast more carefully in terms of limits, this is the Law of Large Numbers

(see Rice, p 178)

We will return to $\bar{X}$ after a short excursion

Summary

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

$$E(h(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) f(x,y)\,dx\,dy$$

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

($= 0$ if X, Y independent)

$$Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$$

$$Var(X_1 + \cdots + X_n) = \sum_{i=1}^{n} Var(X_i) + \sum_{i \neq j}\sum Cov(X_i, X_j)$$

Indep. Case

$$Var(X+Y) = Var(X) + Var(Y)$$

$$Var(\sum X_i) = \sum_{i=1}^{n} Var(X_i)$$

$$Var(\bar{X}) = \frac{\sigma^2}{n} \quad \text{if } Var(X_1) = Var(X_2) = \cdots = \sigma^2$$

$$E(\bar{X}) = \mu \quad \text{if } E(X_1) = E(X_2) = \cdots = \mu$$

Now let's digress, for a moment...

Def The r^{th} moment of random variable X is $E(X^r)$. (if $E(X^r)$ exists!)

We have already looked at the first moment, $E(X)$, and the second moment $E(X^2)$.

Def The r^{th} central moment of rand. var X is

$$E\left[(X - E(X))^r\right]$$

1^{st} central moment: zero (the mean)
2^{nd} central moment: variance, etc.

It turns out that there is a great trick for dealing with moments – the moment-generating function (mgf)

$$M(t) = E\left[e^{tX}\right]$$

Why care about the moment-generating function? (we will consider the cont. case)

$$M(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$

$$M'(t) = \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$

$$= \int_{-\infty}^{\infty} x e^{tx} f(x)\,dx$$

For $t = 0$

$$M'(0) = \int_{-\infty}^{\infty} x f(x)\,dx = E(X)$$

$$M''(t) = \frac{d}{dt}\int_{-\infty}^{\infty} x e^{tx} f(x)\,dx$$

$$= \int_{-\infty}^{\infty} x^2 e^{tx} f(x)\,dx$$

$$M''(0) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = E(X^2)$$

without proof, it turns out that if $M(t)$ exists in an open interval containing zero $M^{(r)}(0) = E(X^r)$

Moreover

If the moment-generating function exists for t in an open interval containing zero, it uniquely determines the probability distribution.

So we can work with mgf's if we want to, instead of pdf's or cdf's

$$\text{mgf} \rightleftarrows \text{pdf} \rightleftarrows \text{cdf}$$

The properties of expectations that we already know enable us to deduce important properties of the mgf.

say X has mgf $M_X(t)$ and $Y = a + bX$

$$\begin{aligned} M_Y(t) &= E(e^{tY}) \\ &= E(e^{at+btX}) \\ &= E(e^{at}e^{btX}) \\ &= e^{at}E(e^{btX}) \end{aligned}$$

$$\boxed{M_Y(t) = e^{at} M_X(bt)}$$

Say X and Y are indep. rand. variables with mgf's M_X and M_Y.

Let $Z = X + Y$

$$M_Z(t) = E(e^{tZ})$$
$$= E(e^{tX+tZ})$$
$$= E(e^{tX}e^{tY})$$

because X and Y are indep $\longrightarrow$ $= E(e^{tX})\,E(e^{tY})$

$$\boxed{M_Z(t) = M_X(t)\,M_Y(t)}$$

Example: The Standard Normal Dist.

$$M(t) = E(e^{tX}) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tx}e^{-x^2/2}\,dx$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}+tx}\,dx$$

note that

$$\frac{x^2}{2} - tx = \frac{1}{2}(x^2 - 2tx + t^2) - \frac{t^2}{2}$$
$$= \frac{1}{2}(x-t)^2 - t^2/2$$

so $X \sim N(0,1)$, as we were saying

$$M(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2} + tx} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2 + t^2/2} dx$$

$$e^{t^2/2} \underbrace{\left(\frac{1}{\sqrt{2\pi}}\right) \int_{-\infty}^{\infty} e^{-(x-t)^2/2} dx}_{= 1}$$

(set $u = \cancel{x} - t$ change of vars)

$$\boxed{M(t) = e^{t^2/2}}$$

With this information in hand, consider $X_1, X_2, X_3, \ldots, X_n$ identically and independently distributed random variables with cdf F. Let

$$S_n = \sum_{i=1}^{n} X_i$$

suppose n grows without limit. What happens to S_n?

$X_1, \dots X_n$ iid random vars

$$S_n = \sum_{i=1}^{n} X_i, \quad E(X)=0, \; Var(X) = \sigma^2$$

Set $Z_n = \dfrac{S_n}{\sigma\sqrt{n}}$. Let's look at the mgf of Z_n as n gets larger and larger. Because the X_n are indep,

$$M_{S_n}(t) = [m(t)]^n \qquad \text{(p. 13, these notes)}$$

and

$$M_{Z_n}(t) = \left[m\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

We can expand m in a Taylor series around zero, so

$$M(s) = M(0) + s\,M'(0) + \tfrac{1}{2}s^2 M''(0) + \dots$$

$E(x) = 0$, so $M'(0) = 0$

$Var(x) = \sigma^2$ so $M''(0) = \sigma^2$

remember, we've considering

$$S_n = \sum_{i=1}^{n} X_i \text{ (i.i.d)} \; ; \quad Z_n = \frac{S_n}{\sigma\sqrt{n}}, \quad M_{Z_n}(t) = \left[m\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

---

from the Taylor expansion, we now have

$$M\left(\frac{t}{\sigma\sqrt{n}}\right) = \underset{m(0)}{1} + \underset{m'(0)}{0} + \frac{1}{2}\sigma^2\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \ldots$$

So $$M_{Z_n}(t) = \left(1 + \frac{1}{2}\sigma^2\frac{t^2}{\sigma^2 n} + \ldots\right)^n$$

$$= \left(1 * \left(\frac{t^2}{2}\right)\frac{1}{n} + \ldots\right)^n$$

Drop the higher order terms, so

$$M_{Z_n}(t) \doteq \left(1 + \left(\frac{t^2}{2}\right)\frac{1}{n}\right)^n$$

but $\lim_{n\to\infty}\left(1 + \frac{a}{n}\right)^n = e^a$

unrigorously, we have calculated that as $n \to \infty$, $M_{Z_n}(t) \to e^{t^2/2}$.

Hence the distribution of Z_n tends to the standard normal!

we are considering   The X's are i.i.d.!

$$S_n = \sum_{i=1}^{n} X_i, \quad Z_n = \frac{S_n}{\sigma\sqrt{n}}, \quad M_{Z_n}(t) = \left[M\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

So, we have just shown that if

$$Z = \lim_{n\to\infty} Z_n$$

$$Z \sim N(0,1).$$

We derived (not proved) the

Central Limit Theorem

This is worth seeing, if only to understand a major reason* why the Normal Distribution is so important. A rigorous proof requires showing that the mgf's exist and that their convergence leads to convergence in distribution.

* Not the only one!