Lecture 2

Basic statistics Review

Discrete Random Variables versus Continuous Random Variables.

Discrete RV

Bernoulli (binary) rv

$$Pr(X=1) = p \quad , \quad Pr(X=0) = 1-p$$

where p = probability of success $0 \le p \le 1$.

Example: flip of fair coin ($p = \frac{1}{2}$)

Let $X = 1$ when head occurs

$X = 0$ when tail occurs

Consider the sample space of 3 flips of a fair coin:

$\{H,H,H\}$, $\{H,H,T\}$, $\{H,T,H\}$, $\{T,H,H\}$

$\{H,T,T\}$, $\{T,T,H\}$, $\{T,H,T\}$, $\{T,T,T\}$

There are 8 possible outcomes in the sample space.

Lecture 2

Let S = no. of successes (heads) in 3 coin tosses.

Then $Pr(S = s = 3) = 1/8$

$Pr(S = s = 2) = 3/8$

$Pr(S = s = 1) = 3/8$

$Pr(S = s = 0) = 1/8$

Probability Density Function (pdf)

S is a binomial rv with $N = 3$ trials and the probability of success of $p = 1/2$.

Then from basic probability theory we know

$$Pr(S = s) = \binom{N}{s} p^s (1-p)^{N-s}$$

where $$\binom{N}{s} = \frac{N!}{(N-s)!\,s!}$$

and $N! = N(N-1)(N-2)\cdots 2\cdot 1$ and similarly for $s!$ and $(N-s)!$ with $0! = 1$.

Lecture 2

In our case $N = 3$, $p = 1/2$

$$pr(S = s = 3) = \binom{3}{3}\left(\frac{1}{2}\right)^3\left(1-\frac{1}{2}\right)^0$$

$$= 1\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^0$$

$$= \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

$$pr(S = s = 2) = \binom{3}{2}\left(\frac{1}{2}\right)^2\left(1-\frac{1}{2}\right)^1$$

$$= 3\left(\frac{1}{2}\right)^3 = 3/8$$

$$pr(S = s = 1) = \binom{3}{1}\left(\frac{1}{2}\right)^1\left(1-\frac{1}{2}\right)^2$$

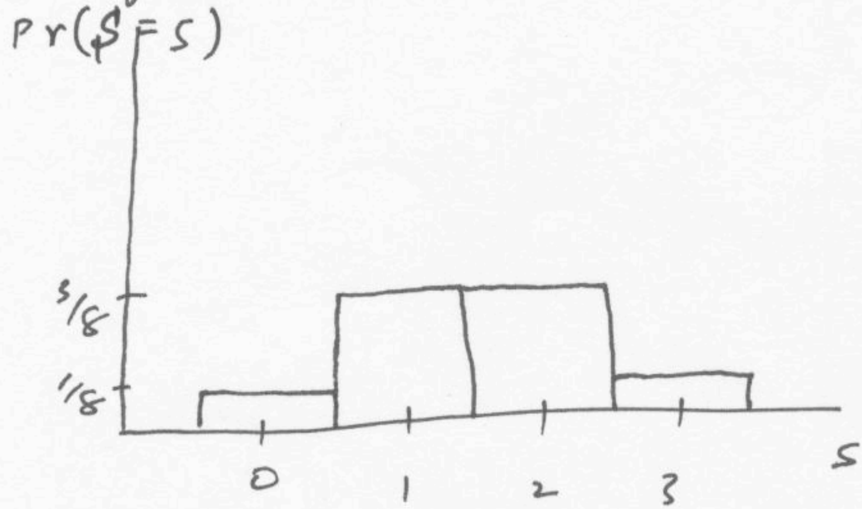
$$= 3\left(\frac{1}{2}\right)^3 = 3/8$$

$$pr(S = s = 0) = \binom{3}{0}\left(\frac{1}{2}\right)^0\left(1-\frac{1}{2}\right)^3$$

$$= 1\left(\frac{1}{2}\right)^3 = 1/8$$

Lecture 2

A graph of the binomial pdf for $N=3$, $p=\frac{1}{2}$.



Notice that the sum of the areas of the rectangles add to one. S is a discrete rv because S takes on only a finite (discrete) number of outcomes.

Cumulative Distribution Function (cdf)

$\Pr(S \le s)$

Lecture 2

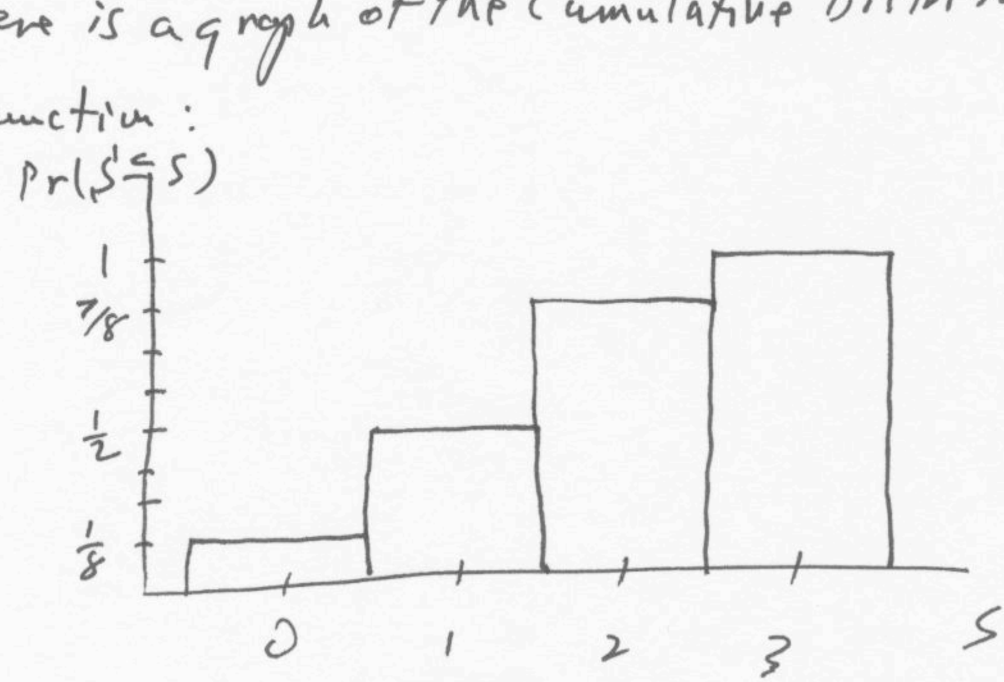
$$Pr(S' \le 0) = \frac{1}{8}$$

$$Pr(S' \le 1) = Pr(S' = 0) + Pr(S' = 1)$$
$$= \frac{1}{8} + \frac{3}{8} = 1/2$$

$$Pr(S' \le 2) = Pr(S' = 0) + Pr(S' = 1) + Pr(S' = 2)$$
$$= \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = 7/8$$

$$Pr(S \le 3) = Pr(S' = 0) + Pr(S' = 1) + Pr(S' = 2)$$
$$+ Pr(S' = 3) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$$

Here is a graph of the Cumulative Distribution Function:



Here the areas don't sum to one but the rectangles eventually grow to have an area of one.

Lecture 2

Let us turn to the derivations of the mean and variance of a discrete rv.

For the discrete rv X with possible outcomes

$x_1, x_2, \cdots, x_r$

The expected value of X, $E(X)$ is calculated as

$$E(X) = \sum_{i=1}^{r} x_i \, Pr(X = x_i)$$

The variance of X is calculated as

$$V(X) = \sum_{i=1}^{r} (x_i - E(X))^2 \cdot Pr(X = x_i).$$

For the three coin toss problem

$$E({}_1S^1) = \sum_{s=0}^{3} s \cdot Pr({}_1S^1 = s)$$

$$= 0 \cdot Pr({}_1S^1 = 0) + 1 \cdot Pr({}_1S^1 = 1) + 2 \cdot Pr({}_1S^1 = 2)$$

$$+ 3 \cdot Pr({}_1S^1 = 3)$$

$$= 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8}$$

$$= \frac{3 + 6 + 3}{8} = 1.5$$

Lecture 2

$$V(S') = \sum_{s=0}^{3} (s-1.5)^2 \cdot Pr(S'=s)$$

$$= (0-1.5)^2 \cdot Pr(S'=0) + (1-1.5)^2 \cdot Pr(S'=1)$$

$$+ (2-1.5)^2 \cdot Pr(S'=2) + (3-1.5)^2 \cdot Pr(S'=3)$$

$$= (-1.5)^2 \cdot \frac{1}{8} + (-0.5)^2 \cdot \frac{3}{8} + (0.5)^2 \cdot \frac{3}{8} + (1.5)^2 \cdot \frac{1}{8}$$

$$= 0.75$$

Continuous Probability Density Function (pdf)

The continuous pdfs are different than the discrete pdfs. In the continuous pdf the probability of the continuous rv taking on the specific value of a point is zero, ie. $Pr(X = x_0) = 0$. Only intervals like $(x_1 < X < x_2)$ have probability. Let $f(x)$ be the pdf of a continuous rv X. Then

(i) $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$

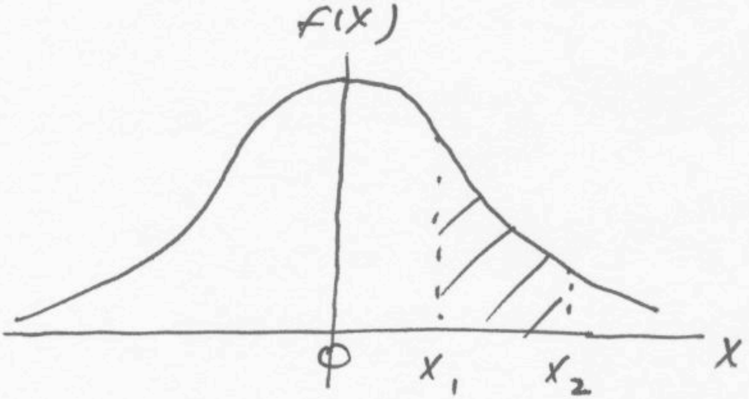
Lecture 2

(ii) $Pr(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)\,dx$

Example: $X \sim N(0,1)$

Then

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$



$$Pr(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)\,dx = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx$$

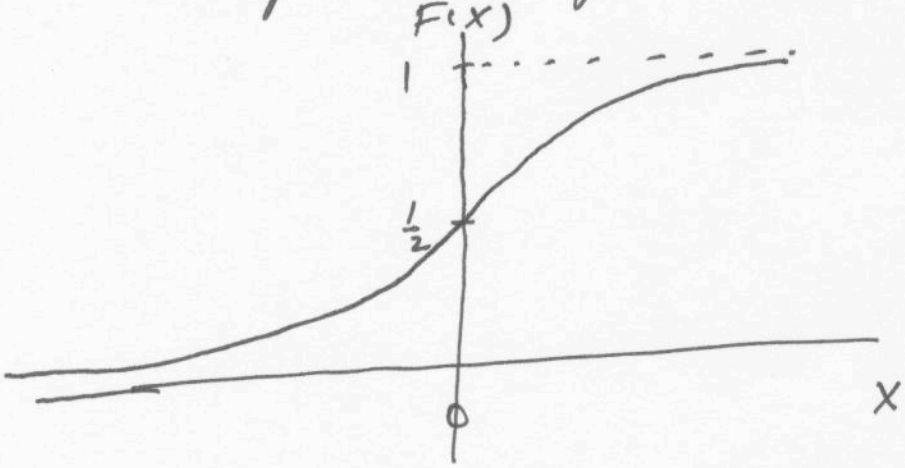
= area under the pdf from x_1 to x_2.

The cumulative distribution function for a continuous rv X is defined as

Lecture 2

$$F(x) = Pr(X \leq x) = \int_{-\infty}^{x} f(x)\,dx$$

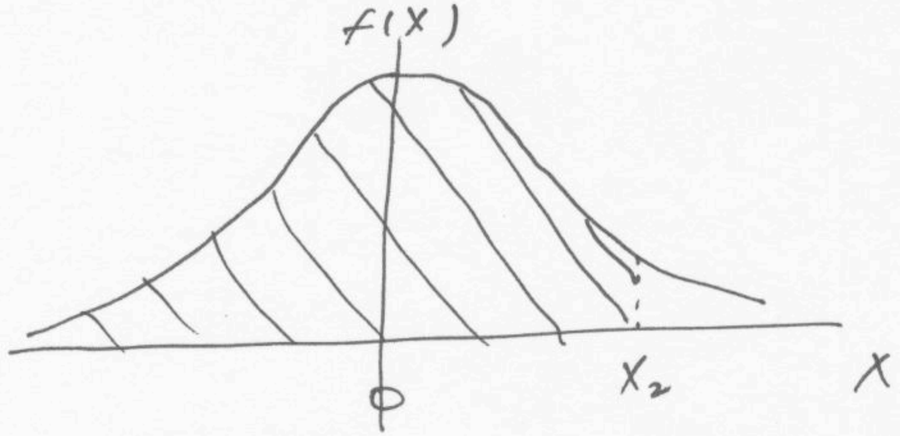
and is represented graphically by (for the $X \sim N(0,1)$)



If $X \sim N(0,1)$ you can calculate

$$Pr(x_1 < X < x_2) = F(x_2) - F(x_1)$$

where $F(x_2)$ looks like



the shaded area under the pdf.

Lecture 2

Example (use your standard normal table in the back of your Wooldridge book). Let $X \sim N(0,1)$

$$pr(1 < X < 2) = F(2) - F(1)$$

$$= 0.9772 - 0.8413$$

$$= 0.1359$$

In a similar vein,

Let $X \sim N(50, 15)$

Calculate

$$pr(45 < X < 55).$$

Let $Z = \frac{X - \mu}{\sigma}$

Then $pr(45 < X < 55) = pr\left(\frac{45-50}{\sqrt{15}} < \frac{X-50}{\sqrt{15}} < \frac{55-50}{\sqrt{15}}\right)$

$$= pr\left(\frac{-5}{\sqrt{15}} < Z < \frac{5}{\sqrt{15}}\right)$$

$$= pr\left(\frac{-5}{3.873} < Z < \frac{5}{3.873}\right) = pr(-1.29 < Z < 1.29)$$

$$= F_Z(1.29) - F_Z(-1.29) = 0.9015 - 0.0985$$

$$= 0.803$$

Lecture 2

For a continuous rv X the population mean is calculated as

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

The variance is calculated as

$$V(X) = \int_{-\infty}^{\infty} (x - E(x))^2 f(x)\,dx.$$

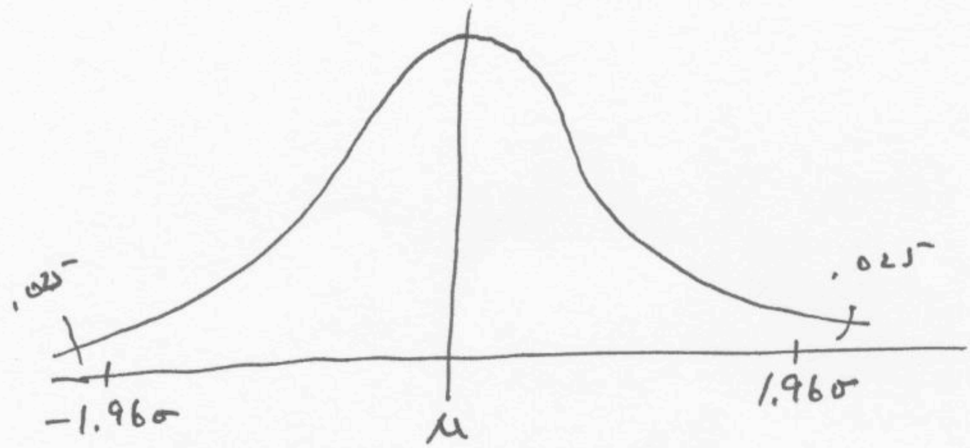
Some popular continuous rv are:

1) $X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

and looks like

Lecture 2



see the next page for a summary of some popular continuous pdfs.

Lecture 2

Continuous Distributions

Distribution	pdf	mean	Variance
Uniform $X \sim U(a,b)$	$f(x) = \frac{1}{b-a}$ for $(a < x < b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal $X \sim N(\mu, \sigma^2)$	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)}{\sigma^2}}$ for $(-\infty < x < \infty)$	μ	σ^2
Exponential	$f(x) = \frac{1}{\beta} e^{-x/\beta}$, $\beta > 0$, $0 \le x < \infty$	β	β^2
Gamma	$f(x) = \left[\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\right] x^{\alpha-1} e^{-x/\beta}$	$\alpha\beta$	$\alpha\beta^2$
Chi-square	$f(\chi^2) = \frac{(\chi^2)^{(\nu/2)-1} e^{-\chi^2/2}}{2^{\nu/2}\Gamma(\nu/2)}$	ν	2ν
Beta	$f(x) = \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right] x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha-\beta)^2(\alpha+\beta+1)}$

t-distribution $\quad T = \frac{Z}{\sqrt{\chi^2/\nu}}$ and as $\nu \to \infty$ $T \to Z$ r.v.

Lecture 2

Continuos Distributions continued

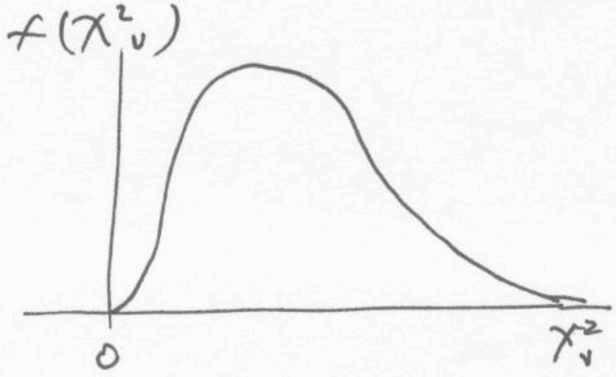
Distribution	pdf	mean	variance
t-distribution	$f(x) = \dfrac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)} \cdot \dfrac{1}{\sqrt{\nu\pi}} \cdot \dfrac{1}{(1+x^2/\nu)^{(\nu+1)/2}}$ $-\infty < x < \infty$	0	$\dfrac{\nu}{\nu-2}$

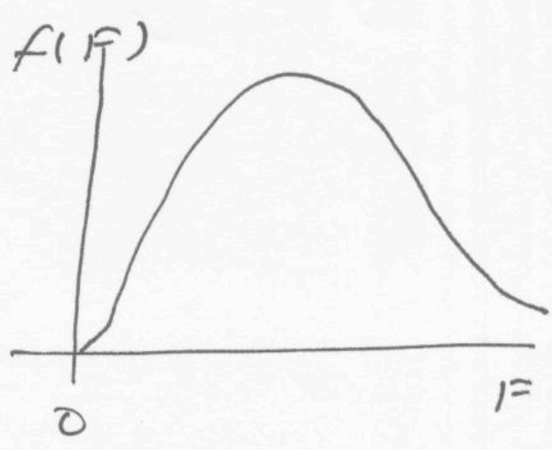
The t-distribution looks like the $N(0,1)$ distribution but has "fatter" tails.

F-distribution
m = numerator df
n = denominator df

$f(x)$ = Beta density with

$$\alpha = \frac{m}{2} \quad \beta = \frac{n}{2}.$$





χ^2 and F look similar.

Lecture 2

Covariance : $E(X-\mu_X)(Y-\mu_Y)$

correlation : $\rho_{XY} = \dfrac{E(X-\mu_X)(Y-\mu_Y)}{\sqrt{Var(X)\cdot Var(Y)}}$

$$-1 \le \rho_{XY} \le 1$$

Properties of Expectations (both discrete and continuous pdfs)

(i) $E(c) = c$, c is constant

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

(iii) $E\sum_{i=1}^{N} a_i X_i = \sum_{i=1}^{N} a_i E(X_i)$

Properties of Variance

(i) $Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y) + 2ab\,cov(X,Y)$

(ii) $var(c) = 0$

Properties of Covariance

(i) $\mathrm{cov}(a_1 X + b_1, a_2 Y + b_2)$

$$= a_1 b_1 \mathrm{cov}(X, Y)$$

(ii) when X and Y are independent

$$\mathrm{cov}(X, Y) = 0.$$