Probability 2

Jacob M. Montgomery

2018

Probability 2

Joint distributions

- Often, we are interested in two or more random variables defined on the same sample space.
  The distribution of these variables is called a **joint distribution**.
- Joint distributions can be made up of any combination of discrete and continuous random variables.

Example

- Suppose we are interested in the outcomes of flipping a coin and rolling a 6-sided die at the same time.
- The sample space for this process contains 12 elements:

$$\{h1, h2, h3, h4, h5, h6, t1, t2, t3, t4, t5, t6\}$$

- We can define two random variables X and Y such that $X = 1$ if heads and $X = 0$ if tails, while Y equals the number on the die.
- We can then make statements about the joint distribution of X and Y.

Joint discret random variables

- If both X and Y are discrete, their joint probability mass function assigns probabilities to each pair of outcomes

$$p(x, y) = \Pr(X = x, Y = y)$$

- Again, $p(x, y) \in [0, 1]$ and $\sum \sum p(x, y) = 1$.

Marginal pmf

- If we are interested in the marginal probability of one of the two variables (ignoring information about the other variable), we can obtain the marginal pmf by summing across the variable that we don't care about:

$$p_X(x) = \sum_i p(x, y_i)$$

Conditional pmf

- We can also calculate the conditional pmf for one variable, holding the other variable fixed.
- Recalling from the previous lecture that $\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$, we can write the conditional pmf as

$$p_{Y|X}(y|x) = \frac{p(x,y)}{p_X(x)}, \quad p_X(x) > 0$$

Joint continuous random variables

- If both X and Y are continuous, their joint probability density function defines their distribution:

$$\Pr((X, Y) \in A) = \iint_A f(x, y) dx dy$$

- Likewise, $f(x, y) \geq 0$ and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$.

Marginal pdf

- Instead of summing, we obtain the marginal probability density function by integrating out one of the variables:

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

Conditional pdf

- Finally, we can write the conditional pdf as

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}, \quad f_X(x) > 0$$

Expectations and moments

- We often want to summarize some characteristics of the distribution of a random variable.
- The most important summary is the expectation (or expected value, or mean), in which the possible values of a random variable are weighted by their probabilities.

Expecation of Discrete Random Variable

- The expected value of a discrete random variable Y is

$$E(Y) = \sum_{y} yp(y)$$

- In words, it is the weighted average of the possible values y can take on, weighted by the probability that y occurs.
- It is not necessarily the number we would expect Y to take on, but rather the average value of Y after a large number of repetitions of an experiment.

Example

- For a fair die,

$$E(Y) = \sum_{y=1}^{6} yp(y) = \frac{1}{6} \sum_{y=1}^{6} y = 7/2$$

- We would never expect the result of a rolled die to be $7/2$, but that would be the average over a large number of rolls of the die.

Expectation of a Continuous Random Variable

- The expected value of a continuous random variable is similar in concept to that of the discrete random variable, except that instead of summing using probabilities as weights, we integrate using the density to weight.
- Hence, the expected value of the continuous variable Y is defined by

$$E(Y) = \int_{-\infty}^{\infty} y f(y) dy$$

Example

Find $E(Y)$ for $f(y) = \frac{1}{1.5}, \quad 0 < y < 1.5.$

$$E(Y) = \int_0^{1.5} \frac{1}{1.5} y dy = \left. \frac{1}{3} y^2 \right|_0^{1.5} = .75$$

Expected Value of any probability function

1. Discrete: $E[g(Y)] = \sum_{y} g(y)p(y)$
2. Continuous: $E[g(Y)] = \int_{-\infty}^{\infty} g(y)f(y)dy$

Other properties of expected values

- $E(c) = c$

Other properties of expected values

- $E(c) = c$
- $E[E[Y]] = E[Y]$ (because the expected value of a random variable is a constant)

Other properties of expected values

- $E(c) = c$
- $E[E[Y]] = E[Y]$ (because the expected value of a random variable is a constant)
- $E[cg(Y)] = cE[g(Y)]$

Other properties of expected values

- $E(c) = c$
- $E[E[Y]] = E[Y]$ (because the expected value of a random variable is a constant)
- $E[cg(Y)] = cE[g(Y)]$
- $E[g(Y_1) + \cdots + g(Y_n)] = E[g(Y_1)] + \cdots + E[g(Y_n)]$
- $E(X|X) = X$
- $E_y(E_x(Y|X)) = E(Y)$

Other properties of expected values

- $E(c) = c$
- $E[E[Y]] = E[Y]$ (because the expected value of a random variable is a constant)
- $E[cg(Y)] = cE[g(Y)]$
- $E[g(Y_1) + \cdots + g(Y_n)] = E[g(Y_1)] + \cdots + E[g(Y_n)]$
- $E(X|X) = X$
- $E_y(E_x(Y|X)) = E(Y)$
- If $X \geq Y$, then $E(X) \geq E(Y)$ with probability 1

Other properties of expected values

- $E(c) = c$
- $E[E[Y]] = E[Y]$ (because the expected value of a random variable is a constant)
- $E[cg(Y)] = cE[g(Y)]$
- $E[g(Y_1) + \cdots + g(Y_n)] = E[g(Y_1)] + \cdots + E[g(Y_n)]$
- $E(X|X) = X$
- $E_y(E_x(Y|X)) = E(Y)$
- If $X \geq Y$, then $E(X) \geq E(Y)$ with probability 1
- $X \perp\!\!\!\perp Y$, then $E(XY) = E(X)E(Y)$

Variance

- We can also look at other summaries of the distribution, which build on the idea of taking expectations.

Variance

- We can also look at other summaries of the distribution, which build on the idea of taking expectations.
- Variance tells us about the “spread" of the distribution; it is the expected value of the squared deviations from the mean of the distribution.

Variance

- We can also look at other summaries of the distribution, which build on the idea of taking expectations.
- Variance tells us about the “spread" of the distribution; it is the expected value of the squared deviations from the mean of the distribution.
- The standard deviation is simply the square root of the variance.

1. Variance: $\sigma^2 = \mathrm{Var}(Y) = E[(Y - E(Y))^2] = E(Y^2) - [E(Y)]^2$
2. Standard Deviation: $\sigma = \sqrt{\mathrm{Var}(Y)}$

Properties of variance

- $Var(X) = E(X - EX)^2$

Properties of variance

- $Var(X) = E(X - EX)^2$
- $Var(X) = E(X^2) - (EX)^2$

Properties of variance

- $Var(X) = E(X - EX)^2$
- $Var(X) = E(X^2) - (EX)^2$
- $Var(c) = 0$

Properties of variance

- $Var(X) = E(X - EX)^2$
- $Var(X) = E(X^2) - (EX)^2$
- $Var(c) = 0$
- $Var(Y|X) = E(Y^2|X) - (E(Y|X))^2$

Properties of variance

- $Var(X) = E(X - EX)^2$
- $Var(X) = E(X^2) - (EX)^2$
- $Var(c) = 0$
- $Var(Y|X) = E(Y^2|X) - (E(Y|X))^2$
- If $X \perp\!\!\!\perp Y$, then $Var(X + Y) = Var(X) + Var(Y)$

Properties of variance

- $Var(X) = E(X - EX)^2$
- $Var(X) = E(X^2) - (EX)^2$
- $Var(c) = 0$
- $Var(Y|X) = E(Y^2|X) - (E(Y|X))^2$
- If $X \perp\!\!\!\perp Y$, then $Var(X + Y) = Var(X) + Var(Y)$
- If $X \perp\!\!\!\perp Y$, then $Var(X - Y) = Var(X) + Var(Y)$

Properties of variance

- $Var(X) = E(X - EX)^2$
- $Var(X) = E(X^2) - (EX)^2$
- $Var(c) = 0$
- $Var(Y|X) = E(Y^2|X) - (E(Y|X))^2$
- If $X \perp\!\!\!\perp Y$, then $Var(X + Y) = Var(X) + Var(Y)$
- If $X \perp\!\!\!\perp Y$, then $Var(X - Y) = Var(X) + Var(Y)$
- Total variation: $Var(Y) = E_x(Var_y(Y|X)) + Var_x(E_y(Y|X))$

Conditional summaries

- The *conditional expection* of Y given X is

$$E(Y|X) = \sum_{i=1}^{k} y_i p(y_i|x)$$

$$E(Y|X) = \int_{-\infty}^{\infty} y_i p(y_i|x)$$

Conditional summaries

- The *conditional expection* of Y given X is

$$E(Y|X) = \sum_{i=1}^{k} y_i p(y_i|x)$$

$$E(Y|X) = \int_{-\infty}^{\infty} y_i p(y_i|x)$$

- Regression is estimating a conditional expectation function.

Conditional summaries

- The *conditional expection* of Y given X is

$$E(Y|X) = \sum_{i=1}^{k} y_i p(y_i|x)$$

$$E(Y|X) = \int_{-\infty}^{\infty} y_i p(y_i|x)$$

- Regression is estimating a conditional expectation function.
- The *conditional variance* of Y given X is

$$Var(Y|X) = E[(Y - E(Y|X))^2|X]$$

Conditional summaries

- The *conditional expection* of Y given X is

$$E(Y|X) = \sum_{i=1}^{k} y_i p(y_i|x)$$

$$E(Y|X) = \int_{-\infty}^{\infty} y_i p(y_i|x)$$

- Regression is estimating a conditional expectation function.
- The *conditional variance* of Y given X is

$$Var(Y|X) = E[(Y - E(Y|X))^2|X]$$

- Regression estimates employ conditional variance assumptions

Conditional variation: Discrete example

Compute $E(Y|X=2)$ and $Var(Y|X=2)$ for the following data:

Conditional variation: Discrete example

Compute $E(Y|X=2)$ and $Var(Y|X=2)$ for the following data:

		X			
		-2	**0**	**2**	**3**
Y	**3**	0.27	0.08	0.16	0
	6	0	0.04	0.10	0.35

$p((X=-2)\cap(Y=3)) = 0.27$, $p((X=3)\cap(Y=6)) = 0.35$, etc.

$$E(Y|X=2) = \sum_{y} y \cdot p(Y=y|X=2)$$

$$\begin{aligned} E(Y|X=2) &= \sum_y y \cdot p(Y=y|X=2) \\ &= 3p(Y=3|X=2) + 6p(Y=6|X=2) \end{aligned}$$

$$
\begin{aligned}
E(Y|X=2) &= \sum_y y \cdot p(Y=y|X=2) \\
&= 3p(Y=3|X=2) + 6p(Y=6|X=2) \\
&= 3(0.16/0.26) + 6(0.10/0.26)
\end{aligned}
$$

$$\begin{aligned} E(Y|X=2) &= \sum_y y \cdot p(Y=y|X=2) \\ &= 3p(Y=3|X=2) + 6p(Y=6|X=2) \\ &= 3(0.16/0.26) + 6(0.10/0.26) \\ &= 4.15 \end{aligned}$$

$$
\begin{aligned}
E(Y|X=2) &= \sum_y y \cdot p(Y=y|X=2) \\
&= 3p(Y=3|X=2) + 6p(Y=6|X=2) \\
&= 3(0.16/0.26) + 6(0.10/0.26) \\
&= 4.15
\end{aligned}
$$

Note: $p(Y=3|X=2) = \frac{p(Y=3,X=2)}{p(X=2)} = 0.16/0.26$, and $p(Y=6|X=2) = \frac{p(Y=6,X=2)}{p(X=2)} = 0.10/0.26$. So,

$$\begin{aligned} E(Y|X=2) &= \sum_y y \cdot p(Y=y|X=2) \\ &= 3p(Y=3|X=2) + 6p(Y=6|X=2) \\ &= 3(0.16/0.26) + 6(0.10/0.26) \\ &= 4.15 \end{aligned}$$

Note: $p(Y=3|X=2) = \frac{p(Y=3,X=2)}{p(X=2)} = 0.16/0.26$, and $p(Y=6|X=2) = \frac{p(Y=6,X=2)}{p(X=2)} = 0.10/0.26$. So,

$$Var(Y|X=2) = \sum_y (Y - E(Y|X=2))^2 p(Y|X=2)$$

$$\begin{aligned} E(Y|X=2) &= \sum_y y \cdot p(Y=y|X=2) \\ &= 3p(Y=3|X=2) + 6p(Y=6|X=2) \\ &= 3(0.16/0.26) + 6(0.10/0.26) \\ &= 4.15 \end{aligned}$$

Note: $p(Y=3|X=2) = \frac{p(Y=3,X=2)}{p(X=2)} = 0.16/0.26$, and $p(Y=6|X=2) = \frac{p(Y=6,X=2)}{p(X=2)} = 0.10/0.26$. So,

$$\begin{aligned} Var(Y|X=2) &= \sum_y (Y - E(Y|X=2))^2 p(Y|X=2) \\ &= (3-4.15)^2(0.16/0.26) + (6-4.15)^2(0.10/0.26) \end{aligned}$$

$$
\begin{aligned}
E(Y|X=2) &= \sum_y y \cdot p(Y=y|X=2) \\
&= 3p(Y=3|X=2) + 6p(Y=6|X=2) \\
&= 3(0.16/0.26) + 6(0.10/0.26) \\
&= 4.15
\end{aligned}
$$

Note: $p(Y=3|X=2) = \frac{p(Y=3,X=2)}{p(X=2)} = 0.16/0.26$, and $p(Y=6|X=2) = \frac{p(Y=6,X=2)}{p(X=2)} = 0.10/0.26$. So,

$$
\begin{aligned}
Var(Y|X=2) &= \sum_y (Y - E(Y|X=2))^2 p(Y|X=2) \\
&= (3-4.15)^2(0.16/0.26) + (6-4.15)^2(0.10/0.26) \\
&= 2.13
\end{aligned}
$$

Continuous case

$$E(Y|X) \;=\; \int_{-\infty}^{\infty} y_i p(y_i|x)$$

But, can't calculate terms of $p(y_i|x)$ as in discrete case. Need integral.

Continuous case

$$E(Y|X) = \int_{-\infty}^{\infty} y_i p(y_i|x)$$

But, can't calculate terms of $p(y_i|x)$ as in discrete case. Need integral. From

$$P(A \cap B) = P(A)P(B|A)$$

Continuous case

$$E(Y|X) = \int_{-\infty}^{\infty} y_i p(y_i|x)$$

But, can't calculate terms of $p(y_i|x)$ as in discrete case. Need integral. From

$$\begin{aligned} P(A \cap B) &= P(A)P(B|A) \\ P(B|A) &= \frac{P(A \cap B)}{P(A)} \end{aligned}$$

Continuous case

$$E(Y|X) = \int_{-\infty}^{\infty} y_i p(y_i|x)$$

But, can't calculate terms of $p(y_i|x)$ as in discrete case. Need integral. From

$$\begin{aligned} P(A \cap B) &= P(A)P(B|A) \\ P(B|A) &= \frac{P(A \cap B)}{P(A)} \\ \text{conditional} &= \frac{\text{joint}}{\text{marginal}} \end{aligned}$$

Given a joint distribution of X and Y of $p(x, y)$,

$$p(y|x) \quad = \quad \frac{p(x, y)}{p(x)}$$

Given a joint distribution of X and Y of $p(x, y)$,

$$p(y|x) \quad = \quad \frac{p(x, y)}{p(x)}$$

To get $p(x)$, integrate joint dist'n over all $\{other\}$ *dimensions.*

Inequalities in expectation

- Markov:
- Chebychev:
- Cauchy-Schwarz: $E(|XY|) \leq \sqrt{EX^2 EY^2}$
- Jensen: If $f(X)$ concave (down), then $E(f(X)) \leq f(EX)$

Inequalities in expectation

- Markov:
- Chebychev:
- Cauchy-Schwarz: $E(|XY|) \leq \sqrt{EX^2EY^2}$
- Jensen: If $f(X)$ concave (down), then $E(f(X)) \leq f(EX)$

- Minkowski:
- H"older:
- Liapounov:
- Cramer-Rao:
- Berge:

Equality of Random Variables

Random variables X and Y are

- Equal in distribution iff

$$p(X \leq x) = p(Y \leq x) \quad \forall x$$

Equality of Random Variables

Random variables X and Y are

- Equal in distribution iff

$$p(X \leq x) = p(Y \leq x) \quad \forall x$$

- Equal in mean iff

$$E(|X - Y|) = 0$$

Equality of Random Variables

Random variables X and Y are

- Equal in distribution iff

$$p(X \leq x) = p(Y \leq x) \quad \forall x$$

- Equal in mean iff

$$E(|X - Y|) = 0$$

- Equal in p-th mean iff

$$E(|X - Y|^p) = 0$$

Covariance

- The *covariance* of two random variables is

$$Cov(X, Y) = E[(X - EX)(Y - EY)]$$

Covariance

- The *covariance* of two random variables is

$$\begin{aligned} Cov(X, Y) &= E[(X - EX)(Y - EY)] \\ &= \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \end{aligned}$$

Covariance

- The *covariance* of two random variables is

$$\begin{aligned} Cov(X, Y) &= E[(X - EX)(Y - EY)] \\ &= \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \end{aligned}$$

- Think about XY plot

- Note that

$$Cov(X, Y) \;=\; E[(X - EX)(Y - EY)]$$

- Note that

$$\begin{aligned} Cov(X, Y) &= E[(X - EX)(Y - EY)] \\ &= E[XY - E(YEX) - E(XEY) + E(EXEY)] \end{aligned}$$

- Note that

$$
\begin{aligned}
Cov(X, Y) &= E[(X - EX)(Y - EY)] \\
&= E[XY - E(YEX) - E(XEY) + E(EXEY)] \\
&= E[XY - EXEY - EYEX + E(EXEY)]
\end{aligned}
$$

- Note that

$$\begin{aligned}
Cov(X,Y) &= E[(X-EX)(Y-EY)] \\
&= E[XY - E(YEX) - E(XEY) + E(EXEY)] \\
&= E[XY - EXEY - EYEX + E(EXEY)] \\
&= E[XY - EXEY - EYEX + EXEY]
\end{aligned}$$

- Note that

$$
\begin{aligned}
Cov(X,Y) &= E[(X-EX)(Y-EY)] \\
&= E[XY - E(YEX) - E(XEY) + E(EXEY)] \\
&= E[XY - EXEY - EYEX + E(EXEY)] \\
&= E[XY - EXEY - EYEX + EXEY] \\
&= E[XY - EXEY]
\end{aligned}
$$

- Note that

$$
\begin{aligned}
Cov(X,Y) &= E[(X-EX)(Y-EY)] \\
&= E[XY - E(YEX) - E(XEY) + E(EXEY)] \\
&= E[XY - EXEY - EYEX + E(EXEY)] \\
&= E[XY - EXEY - EYEX + EXEY] \\
&= E[XY - EXEY] \\
&= E(XY) - EXEY
\end{aligned}
$$

- Note that

$$
\begin{aligned}
Cov(X,Y) &= E[(X-EX)(Y-EY)] \\
&= E[XY - E(YEX) - E(XEY) + E(EXEY)] \\
&= E[XY - EXEY - EYEX + E(EXEY)] \\
&= E[XY - EXEY - EYEX + EXEY] \\
&= E[XY - EXEY] \\
&= E(XY) - EXEY
\end{aligned}
$$

- and

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

$$Cov(X, X) = E[(X - EX)(X - EX)]$$

$$\begin{aligned} Cov(X, X) &= E[(X - EX)(X - EX)] \\ &= E[(X - EX)^2] \end{aligned}$$

$$
\begin{aligned}
Cov(X, X) &= E[(X - EX)(X - EX)] \\
&= E[(X - EX)^2] \\
&= Var(X)
\end{aligned}
$$

$$
\begin{aligned}
Cov(X, X) &= E[(X - EX)(X - EX)] \\
&= E[(X - EX)^2] \\
&= Var(X) \\
&= E[X^2 - 2XEX + (EX)^2]
\end{aligned}
$$

$$
\begin{aligned}
Cov(X, X) &= E[(X - EX)(X - EX)] \\
&= E[(X - EX)^2] \\
&= Var(X) \\
&= E[X^2 - 2XEX + (EX)^2] \\
&= E[X^2] - 2E(XEX) + E[(EX)^2]
\end{aligned}
$$

$$
\begin{aligned}
Cov(X,X) &= E[(X-EX)(X-EX)] \\
&= E[(X-EX)^2] \\
&= Var(X) \\
&= E[X^2 - 2XEX + (EX)^2] \\
&= E[X^2] - 2E(XEX) + E[(EX)^2] \\
&= E[X^2] - 2EXE(X) + (EX)^2
\end{aligned}
$$

$$
\begin{aligned}
Cov(X,X) &= E[(X-EX)(X-EX)] \\
&= E[(X-EX)^2] \\
&= Var(X) \\
&= E[X^2 - 2XEX + (EX)^2] \\
&= E[X^2] - 2E(XEX) + E[(EX)^2] \\
&= E[X^2] - 2EXE(X) + (EX)^2 \\
&= E[X^2] - 2(E(X))^2 + (EX)^2
\end{aligned}
$$

$$
\begin{aligned}
Cov(X,X) &= E[(X-EX)(X-EX)] \\
&= E[(X-EX)^2] \\
&= Var(X) \\
&= E[X^2 - 2XEX + (EX)^2] \\
&= E[X^2] - 2E(XEX) + E[(EX)^2] \\
&= E[X^2] - 2EXE(X) + (EX)^2 \\
&= E[X^2] - 2(E(X))^2 + (EX)^2 \\
&= E(X^2) - (EX)^2
\end{aligned}
$$

Rules of covariance

- If $X \perp\!\!\!\perp Y$, then $Cov(X, Y) = 0$

Rules of covariance

- If $X \perp\!\!\!\perp Y$, then $Cov(X, Y) = 0$ (not iff)

Rules of covariance

- If $X \perp\!\!\!\perp Y$, then $Cov(X, Y) = 0$ (not iff)
- $Cov(a + bX, c + dY) = bdCov(X, Y)$

Rules of covariance

- If $X \perp\!\!\!\perp Y$, then $Cov(X, Y) = 0$ (not iff)
- $Cov(a + bX, c + dY) = bdCov(X, Y)$
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

Moments and moment generating functions

- The k^{th} *moment* of a distribution is

$$E(X^k)$$

Moments and moment generating functions

- The k^{th} *moment* of a distribution is

$$E(X^k)$$

- Thus, first moment $E(X^1) = EX = \bar{x}$

Moments and moment generating functions

- The k^{th} *moment* of a distribution is

$$E(X^k)$$

- Thus, first moment $E(X^1) = EX = \bar{x}$
- The k^{th} *central moment* of a distribution is

$$E[(X - EX)^k]$$

Moments and moment generating functions

- The k^{th} *moment* of a distribution is

$$E(X^k)$$

- Thus, first moment $E(X^1) = EX = \bar{x}$
- The k^{th} *central moment* of a distribution is

$$E[(X - EX)^k]$$

- Thus, second central moment is $E[(X - EX)^2] = Var(X)$

Moments and moment generating functions

- The k^{th} *moment* of a distribution is

$$E(X^k)$$

- Thus, first moment $E(X^1) = EX = \bar{x}$
- The k^{th} *central moment* of a distribution is

$$E[(X - EX)^k]$$

- Thus, second central moment is $E[(X - EX)^2] = Var(X)$
- Third central moment is *skewness* , symmetry

Moments and moment generating functions

- The k^{th} *moment* of a distribution is

$$E(X^k)$$

- Thus, first moment $E(X^1) = EX = \bar{x}$
- The k^{th} *central moment* of a distribution is

$$E[(X - EX)^k]$$

- Thus, second central moment is $E[(X - EX)^2] = Var(X)$
- Third central moment is *skewness* , symmetry
- Fourth central moment is *kurtosis*, tails heavy

Moment generating function

- For any distribution, the moment generating function is defined as:

$$\psi_X(t) = E(e^{tX}) = \int e^{tX} dF(x)$$

How the MGF works

- Take the k^{th} derivative in terms of t
- Set $t = 0$ and solve.
- The answer is the $k^t h$ moment of the distribution
- $$\psi^{(k)}(0) = E(X^k)$$
- This assumes that the integral is is well defined on the open interval around 0.

- Let $X \sim \text{Exp}(1)$. For any $t < 1$

- Let $X \sim \text{Exp}(1)$. For any $t < 1$

$$\begin{aligned} \psi(x) &= \int_0^\infty e^{tx} e^{-x} dx \\ &= \int_0^\infty e^{(t-1)x} dx \end{aligned}$$

- Let $X \sim \text{Exp}(1)$. For any $t < 1$

$$\begin{aligned}
\psi(x) &= \int_0^\infty e^{tx} e^{-x} dx \\
&= \int_0^\infty e^{(t-1)x} dx \\
&= \frac{1}{1-t}
\end{aligned}$$

- Let $X \sim \mathrm{Exp}(1)$. For any $t < 1$

$$
\begin{aligned}
\psi(x) &= \int_0^\infty e^{tx} e^{-x} dx \\
&= \int_0^\infty e^{(t-1)x} dx \\
&= \frac{1}{1-t}
\end{aligned}
$$

- So long as $t < 1$ this is the MGF.

Get involved

- Find $\psi'(0)$
- Find $\psi''(0)$
- Find the first and second central moments

Properties of the MGF

- If $X_1, \ldots, X_n$ are independent and $Y = \sum_i X_i$, then $\psi_Y(t) = \prod_i \psi_i(t)$ where ψ_i is the MGF of X_i
- Let X and Y be random variables. If $\psi_X(t) = \psi_Y(t)$ for all t in an open interval around 0, then X and Y are equal in distribution.
- Pg. 58 in Wasserman lists important moment generating functions.

Random variables in the limit

Aymptotics

- Many times we are interested in the statistical properties of a random variable *in the limit*.
- That is, we want to understand whether/how a random variable will converge as our sample size grows towards infinity.
- For some forms of inference, asymptotic behaviors are essential. For others, they are not.
- But all forms of inference we need asymptotics to evaluate the quality of our estimates.

Types of convergence

Let $X_1, X_2, \ldots$ be a sequence of random variables and let X be some other random variable. Let F_n denote the CDF of X_n and let F denote the CDF of X.

Types of convergence

Let $X_1, X_2, \ldots$ be a sequence of random variables and let X be some other random variable. Let F_n denote the CDF of X_n and let F denote the CDF of X.

1. We say that X_n **converges in probability** to X if for every $\epsilon > 0$

$$P(|X_n - X| > \epsilon) \to 0$$

Types of convergence

Let $X_1, X_2, \ldots$ be a sequence of random variables and let X be some other random variable. Let F_n denote the CDF of X_n and let F denote the CDF of X.

1. We say that X_n **converges in probability** to X if for every $\epsilon > 0$

$$P(|X_n - X| > \epsilon) \to 0$$

2. We say that X_n **converges in distribution** to X, if

$$\lim_{n\to\infty} F_n(t) = F(t)$$

at all t for which F is continuous.

3. We say that X_n **converges in** L_2 to X, if

$$E(X_n - X)^2 \to 0$$

as $n \to \infty$.

Key relationships between types of convergence

1. L_2 convergence implies convergence in probability.

Key relationships between types of convergence

1. L_2 convergence implies convergence in probability.
2. Convergence in probability implies convergence in distribution.

Key relationships between types of convergence

1. L_2 convergence implies convergence in probability.
2. Convergence in probability implies convergence in distribution.

- But note that we cannot reverse these (except in the case that there is a point mass involved.)
- Not further, that some of these convergences hold under transformations (See Theorem 5.5)

Weak law of large numbers

- Says that the the mean of a large sample is close to the mean of the distribution.
- It does not mean that the mean will be equal, but rather that when n is large the distribution of $\bar{X}_n$ will be tightly bound around μ.
- Then the weak law of large numbers states that if $X_1, \ldots, X_n$ are iid, then $\bar{X}_n$ converges in probability to μ.

Proof (fill in the missing parts)

$$P(|\bar{X}_n - \mu| > \epsilon)$$

Proof (fill in the missing parts)

$$P(|\bar{X}_n - \mu| > \epsilon) \leq \frac{\mathrm{Var}(\bar{X}_n)}{\epsilon^2}$$

$$= \frac{\sigma^2}{n\epsilon^2}$$

$$\lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0$$

Example 5.7 in Wasserman

Suppose $X_1, X_2, \ldots X_n$ are results from a fair coin toss. How many times would we have to flip the coin in order be sure that $P(.4 \leq \bar{X}_n \leq .6) \geq .7$?

The mighty central limit theorem

- The CLT helps us approximate probability statements about $\bar{X}_n$
- Suppose that $X_1, X_2, \ldots, X_n$ are *iid* with mean μ and variance σ^2.
- The CLT states that $\bar{X}_n$ has a distribution which is *approximately* Normal with mean μ and variance σ^2/n.
- We assume nothing about the distribution other than the existence of the mean and variance.

- More formally, let $Z_n \equiv \frac{\bar{X}_n - \mu}{\sqrt{\mathrm{Var}(\bar{X}_n)}}$ and Z be the standard normal distribution $N(0, 1)$. Then Z_n converges to Z in distribution.
- Alternatively,
$$\lim_{n \to \infty} P(Z_n \leq z) = \Phi(z)$$

This can be written a couple of different ways (see page 77 in Wasserman). A very common presentation is:

$$\bar{X}_n \approx N(\mu, \sigma^2/n)$$

Example

A recent poll of 698 decided voters in Pennsylvania showed 341 preferred Donald Trump and 357 preferred Hillary Clinton. Let π be the population proportion of decided Pennsylvania voters who prefer Trump. Use the central limit theorem to find the approximate distribution of the sample proportion $\hat{\pi}$.