Last update: Nov 27

List of theorems or properties:

Properties of a probability measure:

- $0 \leq P(A) \leq 1$
- $P(S) = 1$
- If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$
- $P(A^c) = 1 - P(A)$
- $P(A) = P(A \cap B) + P(A \cap B^c)$
- $B \subseteq A$ then $P(A) = P(B) + P(A \cap B^c)$
- $B \subseteq A$ then $P(B) \leq P(A)$
- $A_1, A_2, \ldots$ a partition of S, then $P(B) = \sum_i P(B \cap A_i)$
- $P(A \cup B) = P(A) + P(B) + P(A \cap B)$
- $P(A_1 \cup A_2 \cup \ldots) \leq P(A_1) + P(A_2) + \ldots$

Combinatorics

- $n!$ is the number of permutations of a set of n elements (How many different arrangements)
- $\binom{n}{k} = \frac{n!}{k!(n-k!)} \rightarrow$ choose k elements from a set of n order doesn't matter
- What if order matters? $\binom{n}{k} \times k!$
- $\frac{n!}{(n-k!)}$ Number of ways to pick k items out of n where order does matter.
  - Is the same as $\binom{n}{k} \times k!$ Choosing without order and then calculating permutations of k items.

Conditional probability

- $P(A|B) = \frac{P(A \cap B)}{P(B)}$ if $P(B) > 0$
- $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$
- $A_1, A_2, \ldots$ a partition of S, then $P(B) = \sum_i P(A_i)P(B|A_i) = \sum_i P(B \cap A_i)$
- $P(A|B) = \frac{P(A)}{P(B)} P(B|A)$

Independence

- A and B are independent if $P(A \cap B) = P(A)P(B)$
- $A_1, A_2, A_3, \ldots$ are independent if $P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2})\ldots P(A_{i_k})$ for any finite subcollection of events.

Continuity

- If $\{A_n\} \nearrow A$ then $\lim_{n\to\infty} P(A_n) = P(A)$
- If $\{A_n\} \searrow A$ then $\lim_{n\to\infty} P(A_n) = P(A)$

Random variables (r.v)

- A function $X : S \to \mathbb{R}$ is a random variable
- The distribution of a random variable is the collection of all of the probabilities
  of the variable being in every possible subset of $\mathbb{R}$.
- A r.v. is discrete if $\sum_{x\in\mathbb{R}} P(X = x) = 1$
- Probability function $p_X(k) := P(X = k)$

Discrete random variables

- $X \sim Bernoulli(\theta)$ if

$$p_X(k) = \theta^k (1-\theta)^{1-k}\ \mathbb{1}_{k\in\{0,1\}}$$

$$p_X(k) = \begin{cases} \theta & \text{if } k = 1 \\ 1-\theta & \text{if } k = 0 \\ 0 & \text{otherwise} \end{cases}$$

- $X \sim Binomial(n, \theta)$ if

$$p_X(k) = \binom{n}{k} \theta^k (1-\theta)^{n-k}\ \mathbb{1}_{k\in\{0,1,\ldots,n\}}$$

- $X \sim Geometric(\theta)$ if

$$p_X(k) = (1-\theta)^k \theta \, \mathbb{1}_{k \in \{0,1,\ldots\}}$$

This one counts # of failures before the first success.
there's a second parametrization of geometric counting the number of trials until first success. Check wikipedia!

$$p_Y(k) = (1-\theta)^{k-1} \theta \, \mathbb{1}_{k \in \{1,2,\ldots\}}$$

- $X \sim Poisson(\lambda)$ if

$$p_x(k) = e^{-\lambda} \frac{\lambda^k}{k!} \, \mathbb{1}_{k \in \{0,1,\ldots\}}$$

Poisson approximation

If n is very large and θ very small then $Binomial(n, \theta)$ is well approximated by $Poisson(\lambda = n\theta)$

- $X \sim NegativeBinomial(r, \theta)$ if

$$p_X(k) = \binom{r-1+k}{k} \theta^r (1-\theta)^k \, \mathbb{1}_{k \in \{0,1,\ldots\}}$$

This counts the number of misses before the r^{th} successwith θ as the probability of success.

- $X \sim Hypergeomtric(N, M, n)$ if

$$p_X(k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \, \mathbb{1}_{\max\{0, n+M-N\} \leq k \leq \min\{n, M\}}$$

Here we have a big population of size N composed b 2 sub-populations:
- The population of interest (of size M)
- and the rest of the population (of size $N-M$)

We're taking a sample of size n and X counts the number of elements, in that sample, that come from the population of interest.

Continuous random variables

- A random variable is continuous if $P(X = k) = 0 \; \forall k$
- A density function is "any" function $f : \mathbb{R} \to \mathbb{R}$ with:
  - $f(x) > 0$
  - $\int_{-\infty}^{\infty} f(x) \; dx = 1$
- Then

$$P(a \leq X \leq b) = \int_a^b f(x) \; dx$$

- $X \sim Uniform(a, b)$ for $a < b$ if

$$f(x) = \frac{1}{b-a} \;\; \mathbb{1}_{x \in (a,b)}$$

- $X \sim Exponential(\lambda)$ if

$$f(x) = \lambda e^{-\lambda x} \;\; \mathbb{1}_{x \in (0,\infty)}$$

- $X \sim Normal(\mu, \sigma^2)$ if

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

Cumulative distribution function

Any random variable X the cumulative Distribution Function (CDF) is the function F_X defined by:

$$F_X(x) = P(X \leq x) \; \forall x \in \mathbb{R}$$

- For X discrete

$$F_X(x) = \sum_{u \leq x} P(X = u)$$

- For X absolutely continuous

$$F_x(x) = \int_{-\infty}^{x} f_x(u) \; du$$

Properties:

- $0 \leq F_X(x) \leq 1 \; \forall x \in \mathbb{R}$
- If $x \leq y$ then $F_X(x) \leq F_X(y)$
- $\lim_{x \to -\infty} F_X(x) = 0$
- $\lim_{x \to +\infty} F_X(x) = 1$
- right-continuous

Some other characteristics of CDFs

- If it is also left-continuous, then is continuous and corresponds to a continuous r.v
- If it is differentiable then it corresponds to an absolutely-continuous r.v. and its derivative is the density function of X
- If it's not left-continuous at x then $P[X = x] > 0$
- If X is discrete, then $F_X(x)$ is piece-wise constant

Change of variable theorem

For absolutely continuous r.v.
Suppose X has a density $f_X(x)$ and $Y = h(x)$, where $h : \mathbb{R} \to \mathbb{R}$ is differentiable and strictly monotone (at least on $\{x : f_X(x) > 0\}$) with inverse function $h^{-1}(y)$.
Then Y is also an absolutely continuous r.v. with density function

$$f_Y(y) = \frac{f_X(h^{-1}(y))}{|h'(h^{-1}(y))|}$$

- If $Z \sim Normal(0, 1)$ and $Y = \mu + \sigma Z$ then, $Y \sim Normal(\mu, \sigma^2)$
- If $Y \sim Normal(\mu, \sigma^2)$ then defining $Z = \frac{Y-\mu}{\sigma}$ we get $Z \sim Normal(0, 1)$

Joint Cumulative Distribution Functions

- Given r.v. X and Y their joint cumulative distribution function is the function $F_{X,Y} : \mathbb{R}^2 \to [0, 1]$ given by:

$$F_{X,Y}(a, b) = P(X \leq a, Y \leq b) = P(X \leq a \;\cap\; Y \leq b)$$

Properties of joint CDFs

- $\lim_{a \to -\infty} F_{X,Y}(a,b) = 0 \qquad \forall b \in \mathbb{R}$
- $\lim_{b \to -\infty} F_{X,Y}(a,b) = 0 \qquad \forall a \in \mathbb{R}$
- $\lim_{a \to \infty, b \to \infty} F_{X,Y}(a,b) = 1$
  Marginal CDFs:
- $\lim_{a \to \infty} F_{X,Y}(a,b) = F_Y(b) \qquad \forall b \in \mathbb{R}$
- $\lim_{b \to \infty} F_{X,Y}(a,b) = F_X(a) \qquad \forall a \in \mathbb{R}$

$P(a < X \leq y, c < Y \leq d) = F_{XY}(b,d) - F_{XY}(a,d) - F_{XY}(a,b) + F_{XY}(a,c)$

Joint probability functions (for discrete r.v.)

- joint probability function is

$$p_{X,Y}(a,b) = P(X = a, Y = b)$$

  Marginals
- $\sum_y p_{X,Y}(x,y) = p_X(x)$
- $\sum_x p_{X,Y}(x,y) = p_Y(y)$

Joint density functions (for continuous r.v.)

X, Y are jointly absolutely continuous if there is a joint density function $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ with $f_{X,Y} > 0$ and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dxdy = 1$
Marginals

- $\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx = f_Y(y)$
- $\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy = f_X(x)$

Conditional distribution for discrete r.v.

- In general

$$p_{X|Y}(a|b) = \frac{P(X = a, Y = b)}{P(Y = b)}$$

$$p_{Y|X}(b|a) = \frac{P(Y = b, X = a)}{P(X = a)}$$

Conditional density for continuous r.v.

- Conditional density of Y given $X = x$ is

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

Independence of random variables

- Two random variables X, Y are said to be independent if the joint probability function (or probability density function) is the product of the marginals.

$$f_{X,Y}(x, y) = f_x(x) f_Y(y) \qquad \forall x, y$$

  Equivalently

$$F_{X,Y}(x, y) = F_x(x) F_Y(y) \qquad \forall x, y$$

Multivariable Change-Of-Variable (continuous)

Now $(Z, W) = h(X, Y)$ where $h : \mathbb{R}^2 \to \mathbb{R}^2$.
$Z = h_1(X, Y)$ and $W = h_2(X, Y)$

Need h to be (two-dimensional) differentiable, and one-to-one (invertible).
Then

$$f_{Z,W}(z, w) = \frac{f_{X,Y}(h^{-1}(z, w))}{|J_h(h^{-1}(z, w))|}$$

Where

$$J_h(x, y) = det \begin{pmatrix} \frac{\partial h_1}{\partial x} & \frac{\partial h_1}{\partial y} \\ \frac{\partial h_2}{\partial x} & \frac{\partial h_2}{\partial y} \end{pmatrix}$$

Expected value

For discrete random variables

$$\mathbb{E}(X) = \sum_{t \in S} tP(X = t)$$

For abs. continuous random variables

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} t f_X(t)\ dt$$

Properties of expected values

If X, Y are random variables and $a, b \in \mathbb{R}$. This properties apply to any type of random variable.

- $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$
- If $X \leq Y$ then $\mathbb{E}(X) \leq \mathbb{E}(Y)$
- If X, Y are independent then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$
- If $Z = g(X)$ then $\mathbb{E}(Z) = \mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(t) f_X(t)\ dt$ and the same applies to discrete r.v.

Variance and standard deviation

The variance of any random variable is defined as

$$Var(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$$

We can also compute it as:

- $Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$
- $Var(aX + b) = a^2 Var(X)$

The standard deviation is defined as:

$$sd(X) = \sqrt{Var(X)}$$

Covariance and correlation

Covariance between 2 random variables X, Y is defined as:

$$Cov(X,Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$$

Note the following

- $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
- $Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$
- If X, Y are independent, then $Cov(X,Y) = 0$
- Covariance is bi-linear
  $Cov(aX + bY, Z) = aCov(X,Z) + bCov(Y,Z)$
  $Cov(X, aY + bZ) = aCov(X,Y) + bCov(X,Z)$

Correlation between 2 random variables X, Y with positive variance is defined as:

$$Corr(X,Y) = \rho_{X,Y} := \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

- $\rho_{X,Y} \in [-1, 1]$

Markov's inequality

If $X \geq 0$ and $a > 0$ then

$$P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

Chebychev's Inequality

Let Y be any random variable with finite mean μ_Y. For any $a > 0$ we have

$$P(|Y - \mu_Y| \geq a) \leq \frac{Var(Y)}{a^2}$$

Convergence of Random Variables

Converge in Probability

A sequence $X_1, X_2, X_3, \ldots$ of random variables converges in probability to another random variable (or constant) Y if: For all □ > 0, limn!1 P(jXn□Y j □ □) = 0.

$$X_n \overset{P}{\to} Y \text{ if} \qquad \forall\epsilon > 0 \lim_{n\to\infty} P(|X_n - Y| < \epsilon) = 1$$

- If $\lim_{n\to\infty} P(X_n \neq Y) = 0$ then $X_n \overset{P}{\to} Y$

Weak law of large numbers (WLLN)

Theorem:
For any sequence of random variables $X_1, X_2, X_3, \ldots$ which are independent, and each have the same mean μ, and each have variance $\leq v$ for some constant $v < \infty$,
if $M_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$, then $M_n \overset{P}{\to} \mu$

- Fact: If $\{X_n\}_{n=1}^{\infty}$ are iid then the WLLN doesn't need $v < \infty$

Convergence almost surely (a.s.) (With Probability 1)

A sequence $X_1, X_2, X_3, \ldots$ of r.v. converges almost surely or converges a.s. or converges with probability 1 to another r.v. Y

$$X_n \overset{a.s.}{\to} Y$$

if $P(Xn \to Y \text{as a sequence}) = 1$,
i.e.

$$P(\lim_{n\to\infty} Xn = Y) = 1$$

Theorem: If $X_n \overset{a.s.}{\to} Y$ then $X_n \overset{P}{\to} Y$. Convergence a.s. implies convergence in probability

Strong Law of Large Numbers (SLLN)

Theorem:
For any sequence of random variables $X_1, X_2, X_3, \ldots$ which are iid with mean μ,
if $M_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$, then $M_n \overset{a.s.}{\to} \mu$

Annex

Geometric series

Let a_0 be a constant and define $a_n = a_0 r^n$ as a geometric series

Define $S_n := \sum_{k=0}^{n} r^k$ the partial sums of the first $\{r^k\}_{k=0}^{n}$
note that $\sum_{k=0}^{n} a_k = \sum_{k=0}^{n} a_0 r^k = a_0 \sum_{k=0}^{n} r^k = a_0 S_n$

Now $S_n = r^0 + r^1 + \dots r^n$ and $rS_n = r^1 + r^2 + \dots r^{n+1}$, then

$S_n - rS_n = r^0 - r^{n+1}$
$S_n(1-r) = r^0 - r^{n+1}$
then

$$S_n = \frac{r^0 - r^{n+1}}{1-r} = \frac{1 - r^{n+1}}{1-r} \qquad \text{if } r \neq 1$$

Now if $r \in (0,1)$

$$\sum_{k=0}^{\infty} r^k = \lim_{n\to\infty} \sum_{k=0}^{n} r^k = \lim_{n\to\infty} S_n = \frac{1}{1-r}$$

A little bit of topology

- Remember that both $\emptyset$ and $\mathbb{R}$ are both open and closed sets.
  Consider ONLY nested intervals
- Class of open intervals (a_n, b_n)
  Closed under arbitrary unions and finite intersections
  This means:
- $\bigcup_{n=1}^{\infty}(a_n, b_n)$ will be an open set (either open interval or union of open intervals)
- $\bigcap_{n=1}^{M}(a_n, b_n)$ will be an open set (either open interval or union of open intervals or $\emptyset$)
- $\bigcap_{n=1}^{\infty}(a_n, b_n)$ will be a closed set (closed interval, singleton, $\emptyset$)
- Class of closed intervals $[c_k, d_k]$
  Closed under arbitrary intersections and finite unions
- $\bigcap_{k=1}^{\infty}[c_k, d_k]$ will be a closed set (closed interval, singleton, $\emptyset$)
- $\bigcup_{k=1}^{\infty}[c_k, d_k]$ most likely will be open or have a side that is open $[c, d)$
- $\bigcap_{k=1}^{M}[c_k, d_k]$ will be a closed set

#probability
#ta