Statistical Computing with R – MATH 6382[1,*] Set 2 (Probability and Statistics)

Tamer Oraby
UTRGV
tamer.oraby@utrgv.edu

[1] Could be found in any decent probability and statistics book.

[*] Last updated May 14, 2017

Probability Space

- Random experiment
- Sample space (S or Ω)
- An event $A, B, \ldots \subset S$
- $\sigma-$algebra $\mathcal{B}$
- Probability measure P
- Probability Space $(S, \mathcal{B}, P)$
- Partitioning events
  $\{A_i \subset S : i = 1, 2, \ldots, k; \cup_{i=1}^{k} A_i = S$ and $A_i \cap A_j = \phi$ for all $i \neq j\}$

Independence and Conditional Probability

- Two events A and B are said to be mutually exclusive if and only if $A \cap B = \phi$ and then $P(A \cap B) = 0$
- Two events A and B are said to be independent if and only if $P(A \cap B) = P(A)P(B)$
- The conditional probability of event A given event B is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- Thus, two events A and B are said to be independent if and only if $P(A|B) = P(A)$

Bayes' Theorem

- Law of Total Probability: for a partition $\{A_i\}_{i=1}^n$, and any event B

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

- Bayes' Theorem: for any j ($j = 1, 2, \ldots, n$)

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^n P(B|A_i)P(A_i)}$$

Single Random Variables

Random Variables

- Consider a probability Space $(S, \mathcal{B}, P)$
- A random variable (or r.v.) $X : S \to \mathbb{R}$ is a measurable function
- If range of a random variable X ($\mathcal{R}_X$) is finitely* or infinitely** countable then X is called discrete random variable otherwise (if uncountable) then it is called continuous

*Has the same cardinality[1] as a set $\{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$.
**Has the same cardinality as $\mathbb{N}$.

[1] Two sets A and B have the same cardinality if there exists a bijective (injective+surjective) function $f : A \to B$.

Single Random Variables

A discrete probability *mass* function (pmf) is f such that

$$0 \leq f(x) \leq 1$$

$$\sum_{x \in \mathcal{S}_X} f(x) = 1$$

where $\mathcal{S}_X = \{x \in \mathcal{R}_X : f(x) > 0\}$ is called support of X
A continuous probability *density* function (pdf) is f s.t.

$$0 \leq f(x)$$

$$\int_{\mathbb{R}} f(x)dx = 1$$

Single Random Variables

The expected value of any function $g(X)$ is
If X is discrete r.v. then

$$\mathbf{E}(g(X)) := \sum_{x \in \mathcal{S}_X} g(x) f(x)$$

If X is continuous r.v. then

$$\mathbf{E}(g(X)) := \int_{\mathbb{R}} g(x) f(x) dx$$

If $g(x) = I(x \in A)$ is the indicator function then

$$\mathbf{E}(g(X)) = P(x \in A)$$

Single Random Variables

The r^{th} moment is $\mathbf{E}(X^r)$ and the first moment is called the mean $\mu(X)$ or $\mu_X := \mathbf{E}(X)$ (if exists)
The variance is

$$\mathbf{V}(X) \text{ or } \mathbf{V}_X := \mathbf{E}((X - \mu_X)^2) = \mathbf{E}(X^2) - \mu_X^2$$

The standard deviation is

$$\sigma_X = \sqrt{\mathbf{V}(X)}$$

The moment generating function

$$\mathbf{M}_X(t) := \mathbf{E}(\exp(tX))$$

for all t where $\mathbf{E}(\exp(tX))$ exits
It generates moments

$$\mathbf{E}(X^r) = \frac{d^r\mathbf{M}_X(t)}{dt^r}\,|_{t=0}$$

for $r = 1, 2, \ldots$

Single Random Variables

The cumulative distribution function (cdf) is

$$F(x) := P(X \leq x) = P(\{\omega \in S : X(\omega) \leq x\})$$

If X is discrete r.v. then

$$F(x) = \sum_{t \in \mathcal{S}_X : t \leq x} f(t), \text{ for all } x \in \mathbb{R}$$

and so

$$f(x) = F(x) - F(x^-)$$

where x^- is such that $x^- < x$ and $x^- \in \mathcal{S}_X$
If X is continuous r.v. then

$$F(x) = \int_{-\infty}^{x} f(t)dt, \text{ for all } x \in \mathbb{R}$$

and so

$$f(x) = \frac{dF(x)}{dx}$$

Single Random Variables

The cdf $F(x)$ is non-decreasing ($F(x_1) \leq F(x_2)$ whenever $x_1 < x_2$), right continuous ($\lim_{\epsilon \downarrow 0} F(x + \epsilon) = F(x)$), and $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

Multiple Random Variables

Multiple Random Variables

The joint pmf or pdf of two random variables X and Y is defined to be $f_{X,Y}(x,y)$ such that $f_{X,Y}(x,y) \geq 0$ and

$$\sum_{x \in \mathcal{S}_X} \sum_{y \in \mathcal{S}_Y} f_{X,Y}(x,y) = 1$$

if X and Y are discrete r.v.'s and

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f_{X,Y}(x,y) dx dy = 1$$

if X and Y are continuous r.v.'s

Multiple Random Variables

A joint cdf is defined as

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$$

which is given by

$$F_{X,Y}(x,y) = \sum_{(t,s)\in\mathcal{S}_{X,Y}: t\leq x \text{ and } s\leq y} f_{X,Y}(t,s)$$

for all $(x,y) \in \mathbb{R}^2$, if X and Y are discrete r.v.'s and so

$$f_{X,Y}(x,y) = F_{X,Y}(x,y) - F_{X,Y}(x^-,y) - F_{X,Y}(x,y^-) + F_{X,Y}(x^-,y^-)$$

and

$$F_{X,Y}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(t,s)dtds$$

for all $(x,y) \in \mathbb{R}^2$, if X and Y are continuous r.v.'s and so

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$$

Multiple Random Variables

Marginal pmf or pdf of X are

$$f_X(x) = \sum_{s \in \mathcal{S}_Y} f_{X,Y}(x,s)$$

and

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,s)ds$$

Marginal pmf or pdf of Y are

$$f_Y(y) = \sum_{t \in \mathcal{S}_X} f_{X,Y}(t,y)$$

and

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(t,y)dt$$

X and Y are said to be independent if and only if

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

Multiple Random Variables

The conditional probability function of X given that $Y = y$ is defined by

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

and the conditional probability function of Y given that $X = x$ is defined by

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

$$P(X \in A|Y = y) = \int_A f_{X|Y=y}(x)dx$$

Multiple Random Variables

X and Y are said to be independent if and only if

$$f_{X|Y=y}(x) = f_X(x)$$

for all x and y
Or if and only if

$$f_{Y|X=x}(y) = f_Y(y)$$

for all x and y

Multiple Random Variables

The expected value of $g(X, Y)$ is

$$\mathbf{E}(g(X,Y)) = \sum_{x \in \mathcal{S}_X} \sum_{y \in \mathcal{S}_Y} g(x,y) f_{X,Y}(x,y)$$

if X and Y are discrete r.v.'s and

$$\mathbf{E}(g(X,Y)) = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x,y) f_{X,Y}(x,y) dxdy$$

if X and Y are continuous r.v.'s

Multiple Random Variables

The $(r, p)^{th}$ moment is

$$\mu_{r,p} := \mathbf{E}(X^r Y^p)$$

and $\mu_X := \mu_{1,0}$ and $\mu_Y := \mu_{0,1}$ (if exist)
The variances are

$$\mathbf{V}(X) := \mu_{2,0} - \mu_X^2$$

and

$$\mathbf{V}(Y) := \mu_{0,2} - \mu_Y^2$$

The standard deviations are

$$\sigma_X = \sqrt{\mathbf{V}(X)}$$

and

$$\sigma_Y = \sqrt{\mathbf{V}(Y)}$$

Multiple Random Variables

Two random variables X and Y are said to be identically distributed if X and Y have the same cumulative probability distribution, $F_X \equiv F_Y$. Thus, $\mu_X = \mu_Y$ and $\sigma_X = \sigma_Y$.

Multiple Random Variables

The co-variance of X and Y is

$$Cov(X, Y) = \mathbf{E}[(X-\mu_X)(Y-\mu_Y)] = \mathbf{E}(XY)-\mathbf{E}(X)\mathbf{E}(Y) = \mu_{1,1}-\mu_{1,0}\,\mu_{0,1}$$

The correlation between X and Y is

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X\,\sigma_Y}$$

By Cauchy-Schwarz inequality $|\rho(X, Y)| \leq 1$
Conditional moments

$$\mathbf{E}(g(X)|Y = y) = \int_{\mathbb{R}} g(x) f_{X|Y=y}(x) dx$$

and

$$\mathbf{E}(g(Y)|X = x) = \int_{\mathbb{R}} g(y) f_{Y|X=x}(y) dy$$

Multiple Random Variables

The joint moment generating function

$$\mathbf{M}_{X,Y}(t,s) := \mathbf{E}(\exp(tX + sY))$$

for all t, s where $\mathbf{E}(\exp(tX + sY))$ exits
It generates moments

$$\mathbf{E}(X^r Y^p) = \frac{\partial^{r+p} M_{X,Y}(t,s)}{\partial t^r \partial s^p} \mid_{t,s=0}$$

for $r, p = 1, 2, \ldots$

Multiple Random Variables

If X and Y are said to be independent then

$$\mathbf{E}(g_1(X)g_2(Y)) = \mathbf{E}(g_1(X))\mathbf{E}(g_2(Y))$$

If X and Y are independent then

$$\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$$

and so

$$Cov(X, Y) = 0$$

(and also $\rho(X, Y) = 0$)
If X and Y are independent then

$$\mathbf{M}_{X+Y}(t) = \mathbf{E}(\exp(t(X+Y))) = \mathbf{M}_X(t)\mathbf{M}_Y(t)$$

Multiple Random Variables

Let X and Y be two r.v.'s and a, b and c are real-valued constants

- $\mathbf{E}(aX + b) = a\,\mathbf{E}(X) + b$
- $\mathbf{V}(aX + b) = a^2\,\mathbf{V}(X)$
- $\mathbf{E}(aX + bY + c) = a\,\mathbf{E}(X) + b\,\mathbf{E}(Y) + c$
- $\mathbf{V}(aX + bY + c) = a^2\,\mathbf{V}(X) + b^2\,\mathbf{V}(Y) + 2a\,b\,Cov(X, Y)$
- $\mathbf{E}(X) = \mathbf{E}(\mathbf{E}(X|Y))$
- $\mathbf{V}(X) = \mathbf{E}(\mathbf{V}(X|Y)) + \mathbf{V}(\mathbf{E}(X|Y))$

Law of Total Probability and Bayes' Theorem

- Law of total probability:
  For discrete r.v.'s

$$f_Y(y) = \sum_t f_{Y|X=t}(y) f_X(t)$$

  For continuous r.v.'s

$$f_Y(y) = \int_{\mathbb{R}} f_{Y|X=t}(y) f_X(t) dt$$

- Bayes' Theorem:
  For discrete r.v.'s

$$f_{X|Y=y}(x) = \frac{f_{Y|X=x}(y) f_X(x)}{\sum_t f_{Y|X=t}(y) f_X(t)}$$

  For continuous r.v.'s

$$f_{X|Y=y}(x) = \frac{f_{Y|X=x}(y) f_X(x)}{\int_{\mathbb{R}} f_{Y|X=t}(y) f_X(t) dt}$$

Multiple Random Variables

It could be extended to several random variables $X_1, X_2, \ldots, X_n$ for $n \geq 1$ with the joint pdf

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) \geq 0$$

and

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = 1$$

If $a_1, a_2, \ldots, a_n$ are real-valued constants

$$\mathbf{E}(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n) = a_1 \, \mathbf{E}(X_1) + a_2 \, \mathbf{E}(X_2) + \cdots + a_n \, \mathbf{E}(X_n)$$

Multiple Random Variables

The r.v.'s $X_1, X_2, \ldots, X_n$ are independent if and only if

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

If $X_1, X_2, \ldots, X_n$ are independent r.v.'s then

$$\mathbf{V}(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n) = a_1^2 \mathbf{V}(X_1) + a_2^2 \mathbf{V}(X_2) + \cdots + a_n^2 \mathbf{V}(X_n)$$

and

$$\mathbf{M}_{a_1 X_1 + a_2 X_2 + \cdots + a_n X_n}(t) = \mathbf{M}_{X_1}(a_1\, t) \cdot \mathbf{M}_{X_2}(a_2\, t) \cdots \mathbf{M}_{X_n}(a_n\, t)$$

If $\{X_i\}_{i=1}^n$ is a family of independent identically distributed random variables (i.i.d.r.v.) then

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_X(x_i)$$

and

$$\mathbf{M}_{X_1 + X_2 + \cdots + X_n}(t) = [\mathbf{M}_X(t)]^n$$

Some Discrete Random Variables

Discrete Uniform Random Variable

X assumes one of the values $x_1, x_2, \ldots, x_n$ with pmf

$$f(x_i) = \frac{1}{n}$$

for $i = 1, 2, \ldots, n$
It has mean

$$\mu_X = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

and variance

$$\mathbf{V}_X = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \bar{x^2} - \bar{x}^2$$

Bernoulli Random Variable

$X \sim \mathit{Bernoulli}(p)$ marks failure by 0 (off) and success by 1 (like off and on OR miss or hit) with pmf

$$f(1) = p$$

(and of course $f(0) = 1 - p$) It can be written as

$$f(x) = p^x(1-p)^{1-x}$$

for $x = 0, 1$
It has mean

$$\mu_X = p$$

and variance

$$\mathbf{V}_X = p(1-p)$$

Binomial Random Variable

$X \sim binom(n, p)$ is the number of successes in n independent trials with probability of success on each trial is $p = P(S)$. It has pmf

$$f(x) = C_x^n p^x (1-p)^{n-x}$$

for $x = 0, 1, 2, \ldots, n$
It has mean

$$\mu_X = np$$

and variance

$$\mathbf{V}_X = np(1-p)$$

$$Bernoulli(p) \text{ is } binom(1, p)$$

If $\{X_i\}_{i=1}^n$ is a family of independent identically distributed random variables (i.i.d.r.v.) with $Bernoulli(p)$ then

$$X = X_1 + X_2 + \cdots + X_n \sim Binom(n, p)$$

Multinomial Random Variable

$X \sim multinom(n, p_1, p_2, \ldots, p_k)$ is the vector $(X_1, X_2, \ldots, X_k)$ such that $\sum_{i=1}^{k} X_i = n$ of numbers of realization of the k partitioning events $(\{A_i\}_{i=1}^{k})$ in n independent trials with probabilities of occurrence on each trial is $p_i = P(A_i)$ for $i = 1, 2, \ldots, k$ and $\sum_{i=1}^{k} p_i = 1$. It has joint pmf

$$f(x_1, x_2, \ldots, x_k) = C^{n}_{x_1, x_2, \ldots, x_k} p_1^{x_1} \cdot p_2^{x_2} \cdots p_k^{x_k}$$

for $x_i = 0, 1, 2, \ldots, n$; $i = 1, 2, \ldots, n$ and $\sum_{i=1}^{k} x_i = n$. It has mean

$$\mu_{X_i} = np_i$$

and variance

$$\mathbf{V}_{X_i} = np_i(1 - p_i)$$

and co-variance

$$Cov(X_i, X_j) = -np_ip_j$$

for $i \neq j$

Multinomial Random Variable

Example: If a single student has probabilities to receive letter grades $A, \ldots, F$ are $P(A) = .1, P(B) = .25, P(C) = .3, P(D) = .25, P(F) = .1$. A realization of the experiment of observing letter grades of randomly selected $n = 6$ students is for instance three students received A, one student received B and two students received F, so $x = (3, 1, 0, 0, 2)$. The probability of that instance to occur is

$$f(3, 1, 0, 0, 2) = \frac{6!}{3!\,1!\,0!\,0!\,2!} .1^3 .25^1 .3^0 .25^0 .1^2$$

Geometric Random Variable

$X \sim geom(p)$ is the number of independent **trials** till the **1st** success occurs, given that probability of success on a single trial is $p = P(S)$. Say in F, F, S, $x = 3$. It has pmf

$$f(x) = p(1-p)^{x-1}$$

for $x = 1, 2, \ldots$.
It has mean

$$\mu_X = \frac{1}{p}$$

and variance

$$\mathbf{V}_X = \frac{1-p}{p^2}$$

The cdf is $F(x) = P(X \le x) = 1 - (1-p)^{\lfloor x \rfloor}$ for all $x \ge 0$ and zero otherwise.

Geometric Random Variable

Memoryless Property

$$
\begin{aligned}
P(X > t + s | X > s) &= \frac{P(X > t + s \text{ and } X > s)}{P(X > s)} \\
&= \frac{P(X > t + s)}{P(X > s)} \\
&= \frac{1 - F(t + s)}{1 - F(s)} \\
&= \frac{(1 - p)^{\lfloor t+s \rfloor}}{(1 - p)^{\lfloor s \rfloor}} \\
&= (1 - p)^{\lfloor t+s \rfloor - \lfloor s \rfloor} \\
&= (1 - p)^{t} \text{ if } t \text{ is an integer} \\
&= P(X > t) \text{ if } t \text{ is an integer}
\end{aligned}
$$

Geometric Random Variable

Another point of view:
$Y \sim geom(p)$ is the number of independent **failures** till the **1st** success occurs, given that probability of success on a single trial is $p = P(S)$. Say in F, F, S, $y = 2$. It has pmf

$$f(y) = p(1-p)^y$$

for $y = 0, 1, 2, \ldots$. Thus, $Y = X - 1$.
It has mean

$$\mu_Y = \mu_X - 1 = \frac{1}{p} - 1 = \frac{1-p}{p}$$

and variance

$$\mathbf{V}_Y = \mathbf{V}_X = \frac{1-p}{p^2}$$

The cdf is $F(y) = P(Y \leq y) = 1 - (1-p)^{\lfloor y+1 \rfloor}$ for all $y \geq 0$ and zero otherwise.

Negative Binomial Random Variable

$X \sim nbinom(r, p)$ is the number of independent **trials** till the $\mathbf{r^{th}}$ success occurs, given that probability of success on a single trial is $p = P(S)$. Say, with $r = 3$, in $F, F, S, F, S, F, F, F, S$, $x = 9$.
It has pmf

$$f(x) = C_{r-1}^{x-1} p^r (1-p)^{x-r}$$

for $x = r, r+1, r+2, \ldots$
It has mean

$$\mu_X = r\frac{1}{p}$$

and variance

$$\mathbf{V}_X = r\frac{1-p}{p^2}$$

$$geom(p) \text{ is } nbinom(1, p)$$

Negative Binomial Random Variable

Another point of view:
$Y \sim nbinom(r, p)$ is the number of independent **failures** till the $\mathbf{r^{th}}$ success occurs, given that probability of success on a single trial is $p = P(S)$. Say, with $r = 3$, in $F, F, S, F, S, F, F, F, S$, $y = 6$. That is, $Y = X - r$.
It has pmf

$$f(y) = C_{r-1}^{y+r-1} p^r (1-p)^y$$

for $y = 0, 1, 2, \ldots$
It has mean

$$\mu_Y = \mu_X - r = r\frac{1}{p} - r = r\frac{1-p}{p}$$

and variance

$$\mathbf{V}_Y = \mathbf{V}_X = r\frac{1-p}{p^2}$$

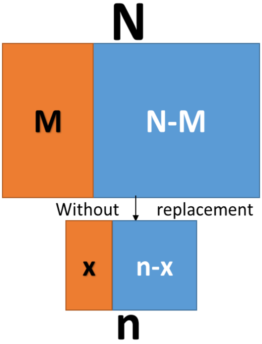
Negative Binomial Random Variable

If $\{X_i\}_{i=1}^{r}$ is a family of independent identically distributed random variables (i.i.d.r.v.) with *geom*(p) then

$$X = X_1 + X_2 + \cdots + X_r \sim nbinom(r, p)$$

Hypergeometric Random Variable

$X \sim hyper(n, M, N)$ is the number of items of a certain type found in a random sample of size n selected without replacement from a population of size N that contains a total of M items of that type.



Hypergeometric Random Variable

It has pmf

$$f(x) = \frac{C_x^M C_{n-x}^{N-M}}{C_n^N}$$

for $x = \max(0, n + M - N), \ldots, \min(n, M)$
It has mean

$$\mu_X = n\frac{M}{N}$$

and variance

$$\mathbf{V}_X = n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1}$$

Hypergeometric Random Variable

Binomial Approximation to Hypergeometric

If $\dfrac{M}{N} \to p$ as $N \to \infty$ while n is fixed then the pmf of $hyper(n, M, N)$ approaches the pmf of $binom(n, p)$ as $N \to \infty$.

Poisson Random Variable

$X \sim pois(\lambda)$ is the number of occurrences of a certain event that is known to happen at a rate of λ per unit space or time.
It has pmf

$$f(x) = \frac{e^{-\lambda}\,\lambda^x}{x!}$$

for $x = 0, 1, 2, \ldots$
It has mean equal to its variance

$$\mu_X = \mathbf{V}_X = \lambda$$

Poisson Random Variable

Poisson Approximation to Binomial

If $np \to \lambda$ as $n \to \infty$ then the pmf of $binom(n, p)$ approaches the pmf of $pois(\lambda)$ as $n \to \infty$.

Some Continuous Random Variables

Continuous Uniform Random Variable

$X \sim unif(a, b)$ so as to $P(X \in (x, x+h)) = P(X \in (y, y+h))$ for all x, y, h such that $(x, x+h)$ and $(y, y+h) \subset (a, b)$
It has a pdf given by

$$f(x) = \frac{1}{b-a} \text{ for } a < x < b$$

where $a < b$, and mean

$$\mu_X = \frac{a+b}{2}$$

variance

$$\mathbf{V}_X = \frac{(b-a)^2}{12}$$

and cdf

$$F(x) = \begin{cases} 0 & \text{if } x < a, \\ \dfrac{x-a}{b-a} & \text{if } a \le x < b, \\ 1 & \text{if } x \ge b. \end{cases}$$

Continuous Uniform Random Variable

If $U \sim unif(0, 1)$, then $X = a + (b - a) * U \sim unif(a, b)$.

Exponential Random Variable

$X \sim exp(\lambda)$ is the simplest way to stochastically model time till success (an event) takes place, e.g., time between transitions made by a Markov process or time between arrivals of customers to an ATM
It has a pdf given by

$$f(x) = \lambda e^{-\lambda x} \text{ for } x \geq 0$$

where the rate $\lambda > 0$ and mean

$$\mu_X = \frac{1}{\lambda}$$

variance

$$\mathbf{V}_X = \frac{1}{\lambda^2}$$

and cdf

$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 - e^{-\lambda x} & \text{if } x \geq 0. \end{cases}$$

Exponential Random Variable

The *survival function* is $S(x) := P(X > x) = 1 - F(x) = e^{-\lambda x}$ whenever $x \geq 0$ and a constant *hazard function* $h(x) := \dfrac{f(x)}{S(x)} = \lambda$

Memoryless Property

$$\begin{aligned}
P(X > t + s | X > s) &= \frac{P(X > t + s \text{ and } X > s)}{P(X > s)} \\
&= \frac{P(X > t + s)}{P(X > s)} \\
&= \frac{S(t + s)}{S(s)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} \\
&= e^{-\lambda t} = P(X > t)
\end{aligned}$$

Exponential Random Variable

Recall that geom and exp distributions have also memoryless property. Also, geom distribution models discrete time till success whereas exp distribution models continuous time till success.
In addition, ...
If $X \sim geom(p)$, let $p = \lambda/n$ and $X = Y/n$ then as $n \to \infty$ the probability distribution of Y approaches the probability distribution of $exp(\lambda)$

Double Exponential (Laplace) Random Variable

$X \sim doublex(\mu, \lambda)$ requires the library "smoothmest" in R. It is two exponential distributions glued back-to-back at a location μ.
It has a pdf given by

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x-\mu|} \text{ for } -\infty < x < \infty$$

where the rate $\lambda > 0$ and mean

$$\mu_X = \mu$$

variance

$$\mathbf{V}_X = \frac{2}{\lambda^2}$$

If $X \sim doublex(\mu, \lambda)$ then $|X - \mu| \sim exp(\lambda)$
If $X, Y \sim exp(\lambda)$ are independent r.v.'s then $X - Y \sim doublex(0, \lambda)$

Gamma Random Variable

$X \sim gamma(r, \lambda)$ to the exponential distribution in continuous r.v.'s is like the negative binomial distribution to the geometric distribution in discrete r.v.'s. It models time as well.
It has a pdf given by

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} \text{ for } x \geq 0$$

where $r, \lambda > 0$ and the special function $\Gamma(r) := \int_0^\infty x^{r-1} e^{-x} dx$ is the gamma function
It has a mean

$$\mu_X = \frac{r}{\lambda}$$

variance

$$\mathbf{V}_X = \frac{r}{\lambda^2}$$

$exp(\lambda)$ is $gamma(1, \lambda)$

Gamma Random Variable

Let r be an integer. If $\{X_i\}_{i=1}^{r}$ is a family of independent identically distributed random variables (i.i.d.r.v.) with $exp(\lambda)$ then

$$X = X_1 + X_2 + \cdots + X_r \sim gamma(r, \lambda)$$

Gamma Random Variable

Remarks about the special function gamma function
$\Gamma(r) := \int_0^\infty x^{r-1}e^{-x}dx$

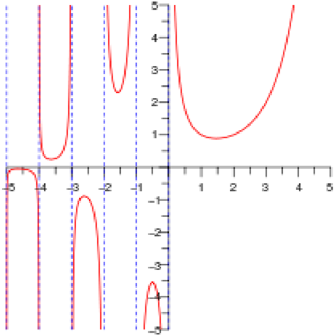
$$\Gamma(r+1) = r\Gamma(r) \text{ for } r \in \mathbb{R} - \{0, -1, -2, \ldots\}$$

and if r is an integer then $\Gamma(r+1) = r!$ with $\Gamma(1) = 0! = 1$

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}$$

$$\Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}$$

$$\Gamma(\frac{5}{2}) = \frac{3}{2}\frac{1}{2}\sqrt{\pi}$$



Chi-square Random Variable

$X \sim chisq(\nu)$ is $gamma(\frac{\nu}{2}, \frac{1}{2})$ where the degrees of freedom $\nu = 1, 2, \ldots$

It has a pdf given by

$$f(x) = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} e^{-x/2} \text{ for } x \geq 0$$

It has a mean

$$\mu_X = \nu$$

variance

$$\mathbf{V}_X = 2\nu$$

Chi-square Random Variable

If $\{X_i\}_{i=1}^n$ is a family of independent random variables with $X_i \sim chisq(\nu_i)$ then

$$\sum_{i=1}^{n} X_i \sim chisq\left(\sum_{i=1}^{n} \nu_i\right)$$

Weibull Random Variable

$X \sim$ *weibull*(κ, λ) is another way to stochastically model time till success (an event) takes place.
It has a pdf given by

$$f(x) = \kappa\lambda^{\kappa}x^{\kappa-1}e^{-(\lambda x)^{\kappa}} \text{ for } x \geq 0$$

where the rate $\kappa, \lambda > 0$ and mean

$$\mu_X = \frac{\Gamma(1+\frac{1}{\kappa})}{\lambda}$$

variance

$$\mathbf{V}_X = \frac{1}{\lambda^2}\left(\Gamma(1+\frac{2}{\kappa}) - \left(\Gamma(1+\frac{1}{\kappa})\right)^2\right)$$

and cdf

$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 - e^{-(\lambda x)^{\kappa}} & \text{if } x \geq 0. \end{cases}$$

Note: *weibull*$(1, \lambda)$ is *exp*(λ)

Beta Random Variable

$X \sim beta(\alpha, \beta)$ is a famous model of a fraction and probability of events, e.g., success.
It has a pdf given by

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1} \text{ for } 0 \leq x \leq 1$$

where $\alpha, \beta > 0$ and the special function
$B(\alpha, \beta) := \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the beta function
It has a mean

$$\mu_X = \frac{\alpha}{\alpha+\beta}$$

variance

$$\mathbf{V}_X = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Beta Random Variable

- *unif*$(0, 1)$ is *beta*$(1, 1)$
- If $X \sim$ *gamma*$(\alpha, 1)$ and $Y \sim$ *gamma*$(\beta, 1)$ are two independent r.v.'s, then

$$\frac{X}{X+Y} \sim \textit{beta}(\alpha, \beta)$$

Normal Random Variable

$X \sim norm(\mu, \sigma)$ is used to model many phenomena and measurements.
It has a pdf given by

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

where $-\infty < \mu < \infty$ and $\sigma > 0$
It has a mean $\mu_X = \mu$ and variance $\mathbf{V}_X = \sigma^2$
$Z = \frac{X-\mu}{\sigma} \sim norm(0, 1)$ the standard normal distribution
The cdf of the standard normal is denoted by

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

Normal Random Variable

- If $X \sim norm(\mu, \sigma)$, and a, b are constant, then

$$aX + b \sim norm(a\mu + b, |a|\,\sigma)$$

- If $\{X_i\}_{i=1}^{n}$ is a family of independent random variables with $X_i \sim norm(\mu_i, \sigma_i)$ then

$$X = \sum_{i=1}^{n} a_i X_i \sim norm\left(\sum_{i=1}^{n} a_i \mu_i, \sqrt{\sum_{i=1}^{n} a_i^2 \sigma_i^2}\right)$$

- If $\{X_i\}_{i=1}^{n}$ is a family of independent random variables with $X_i \sim norm(\mu_i, \sigma_i)$ then

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim norm\left(\frac{1}{n}\sum_{i=1}^{n} \mu_i, \frac{1}{n}\sqrt{\sum_{i=1}^{n} \sigma_i^2}\right)$$

Normal Random Variable

- If $\{X_i\}_{i=1}^n$ is a family of independent identically distributed random variables (i.i.d.r.v.) with $X_i \sim norm(\mu, \sigma)$ then

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim norm(\mu, \frac{\sigma}{\sqrt{n}})$$

- **Chi-square Random Variable (revisited)** If $\{Z_i\}_{i=1}^n$ is a family of independent identically distributed random variables (i.i.d.r.v.) with $Z_i \sim norm(0, 1)$ then

$$\sum_{i=1}^{n} Z_i^2 \sim chisq(n)$$

- Application: $(n-1)S^2/\sigma^2 \sim chisq(n-1)$ where $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is the sample variance.

Log-Normal Random Variable

If $X \sim norm(\mu, \sigma)$ then $e^X \sim lnorm(\mu, \sigma)$. The log-normal r.v. is sometimes used to model time till success or any specific event takes place.
It has a pdf given by

$$f(x) = \frac{1}{x\sqrt{2\pi}\sigma} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} \text{ for } x > 0$$

where $\mu, \sigma > 0$
It has a mean $\mu_X = e^{\mu+\sigma^2/2}$ and variance $\mathbf{V}_X = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$

Student T Random Variable

$X \sim t(\nu)$ is a result of $X = \dfrac{Z}{\sqrt{\chi^2/\nu}}$ where $Z \sim norm(0,1)$ and $\chi^2 \sim chisq(\nu)$ are two independent r.v.'s

It has a pdf given by

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\nu\pi}} \frac{1}{\left(1+\frac{x^2}{\nu}\right)^{\frac{\nu+1}{2}}} \text{ for } -\infty < x < \infty$$

where $\nu = 1, 2, \ldots$

It has a mean $\mu_X = 0$ whenever $\nu > 1$ and variance $\mathbf{V}_X = \dfrac{\nu}{\nu - 2}$ whenever $\nu > 2$

Student T Random Variable

Application: If $\{X_i\}_{i=1}^{n}$ is a family of independent identically distributed random variables (i.i.d.r.v.) with $X_i \sim norm(\mu, \sigma)$ then

$$\bar{X} \sim norm(\mu, \frac{\sigma}{\sqrt{n}})$$

and so, on one hand,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim norm(0, 1)$$

and on the other hand

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim chisq(n-1)$$

$\bar{X}$ and S^2 are independent r.v.'s (proof is out of the scope of the course) then

$$T = \frac{Z}{\sqrt{\chi^2/(n-1)}} = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

Cauchy Random Variable

$X \sim cauchy(\mu, \sigma)$ is a heavy tail distribution (α-stable law with $\alpha = 1$)
It has a pdf given by

$$f(x) = \frac{1}{\pi\sigma} \frac{1}{1 + \left(\frac{x-\mu}{\sigma}\right)^2} \text{ for } -\infty < x < \infty$$

with $\sigma > 0$
The mean, variance and all the moments do not exist. But the cdf is given by

$$F(x) = \frac{1}{2} + \arctan(\frac{x-\mu}{\sigma}) \text{ for } -\infty < x < \infty$$

The Standard Cauchy $cauchy(0, 1)$ is $t(1)$

F Random Variable

$X \sim f(\nu_1, \nu_2)$ is a result of $X = \dfrac{\chi_1^2/\nu_1}{\chi_2^2/\nu_2}$ where $\chi_1^2 \sim chisq(\nu_1)$ and $\chi_2^2 \sim chisq(\nu_2)$ are independent
It has a pdf given by

$$f(x) = \frac{1}{B(\frac{\nu_1}{2}, \frac{\nu_2}{2})} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \frac{x^{\frac{\nu_1}{2}-1}}{\left(1 + \frac{\nu_1}{\nu_2}x\right)^{\frac{\nu_1+\nu_2}{2}}} \text{ for } x > 0$$

where $\nu_1, \nu_2 = 1, 2, \ldots$
It has a mean $\mu_X = \dfrac{\nu_2}{\nu_2 - 2}$ whenever $\nu_2 > 2$ and variance

$\mathbf{V}_X = \dfrac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}$ whenever $\nu_2 > 4$

F Random Variable

Application: If $\{X_i\}_{i=1}^{n}$ and $\{Y_i\}_{i=1}^{m}$ are two independent families of independent identically distributed random variables (i.i.d.r.v.) with $X_i \sim norm(\mu_X, \sigma_X)$ and $Y_j \sim norm(\mu_Y, \sigma_Y)$ then

$$\chi_1^2 = \frac{(n-1)S_X^2}{\sigma_X^2} \sim chisq(n-1)$$

and

$$\chi_2^2 = \frac{(m-1)S_Y^2}{\sigma_Y^2} \sim chisq(m-1)$$

are independent and thus

$$F = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim f(n-1, m-1)$$

Discrete mixture of probability distributions

Let $X_1, X_2, \dots, X_K$ be r.v. with $X_i \sim F_{X_i}(\cdot|\theta_i)$
A discrete-mixture distribution of X is

$$F_X(\cdot|\theta) = p_1 F_{X_1}(\cdot|\theta_1) + p_2 F_{X_2}(\cdot|\theta_2) + \cdots + p_K F_{X_K}(\cdot|\theta_K)$$

where $p_i > 0$ and $p_1 + p_2 + \cdots + p_K = 1$
Example: Flip a coin, if it lands up head use $norm(0, 1)$ otherwise use $norm(0, 2)$, then the resulting r.v. X has a cdf

$$F_X(x) = \frac{1}{2}\Phi(x) + \frac{1}{2}\Phi(\frac{x}{2})$$

Continuous mixture of probability distributions

Let Y be r.v. with $Y \sim F_Y(\cdot|\theta, \lambda_1)$ and $\Theta \sim F_\Theta(\theta|\lambda_2)$
A continuous-mixture distribution of X is

$$F_X(\cdot|\lambda) = \int_{\mathbb{R}} F_Y(\cdot|\theta, \lambda_1)\, f_\Theta(\theta|\lambda_2)\, d\theta$$

where $f_\Theta(\theta|\lambda_2) > 0$ and $\int_{\mathbb{R}} f_\Theta(\theta|\lambda_2)\, d\theta = 1$

Continuous mixture of probability distributions

Example: (Gamma-Poisson mixture)
Let $Y \sim pois(\lambda)$ and $\lambda \sim gamma(r, \beta)$. It is known analytically that the Gamma-Poisson mixture follows $nbiom(r, \frac{\beta}{1+\beta})$, since for each $x = 0, 1, \ldots$

$$\begin{aligned}
f_X(x|r,\beta) &= \int_{\mathbb{R}} f_Y(x|\lambda)\, f_\Lambda(\lambda|r,\beta)\, d\lambda \\
&= \int_0^\infty \frac{\lambda^x}{x!} e^{-\lambda} \frac{\beta^r}{\Gamma(r)} \lambda^{r-1} e^{-\beta\lambda}\, d\lambda \\
&= \frac{\beta^r}{x!\Gamma(r)} \int_0^\infty \lambda^{x+r-1} e^{-(1+\beta)\lambda}\, d\lambda \\
&= \frac{\Gamma(x+r)}{x!\Gamma(r)} \frac{\beta^r}{(1+\beta)^{x+r}} \\
&= \frac{\Gamma(x+r)}{x!\Gamma(r)} \left(\frac{\beta}{1+\beta}\right)^r \left(\frac{1}{1+\beta}\right)^x
\end{aligned}$$

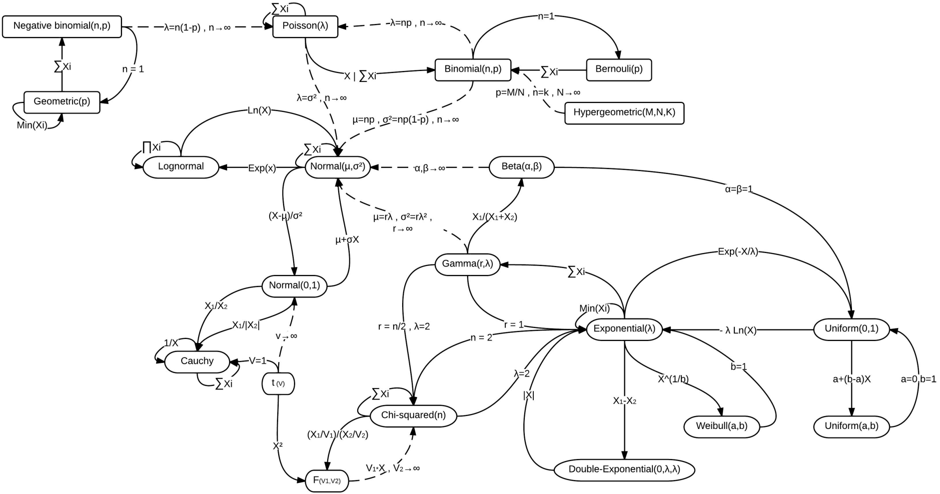
Continuous mixture of probability distributions

Example: (Gamma-Poisson mixture)
If r is an integer, then for each $x = 0, 1, \ldots$

$$\begin{aligned} f_X(x|r,\beta) &= \frac{\Gamma(x+r)}{x!\Gamma(r)} \left(\frac{\beta}{1+\beta}\right)^r \left(\frac{1}{1+\beta}\right)^x \\ &= \frac{(x+r-1)!}{x!(r-1)!} \left(\frac{\beta}{1+\beta}\right)^r \left(\frac{1}{1+\beta}\right)^x \end{aligned}$$

which is $nbiom(r, \frac{\beta}{1+\beta})$. If $r > 0$ a real-number the it is called $polya(r, \mu)$ with $\beta = \frac{\mu}{r}$ which then has mean μ and variance $\mu + \frac{1}{r}\mu^2$. The parameter r (or its reciprocal) is called clustering, aggregation, heterogeneity, or over-dispersion parameter. As $r \to \infty$, $polya(r, \mu)$ approaches $pois(\mu)$.

Map



Source: https://en.wikipedia.org/wiki/Relationships_among_probability_distributions

Multi-variate Normal distribution

Multi-variate Normal distribution

$\mathbf{X} \sim mvnorm(\mu, \Sigma)$ requires the library "mvtnorm" in R, where $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ is a vector of possibly correlated random variables. The mean vector

$$\mu = (\mu_1, \mu_2, \ldots, \mu_d)$$

and the variance-covariance matrix

$$\Sigma = (\sigma_{i,j})_{i,j=1}^{d}$$

is a symmetric positive definite matrix in which $\sigma_{i,j} = Cov(X_i, X_j)$. Note then $\sigma_{i,i} = \sigma_i^2$.

The joint pdf is

$$f(x) = \frac{1}{(2\pi)^{d/2}\sqrt{det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right)$$

for $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$

Each $X_i \sim norm(\mu_i, \sigma_i)$

Bi-variate Normal distribution

$\mathbf{X} \sim mvnorm(\mu, \Sigma)$ where $\mathbf{X} = (X_1, X_2)$ is a vector of possibly correlated random variables. The mean vector

$$\mu = (\mu_1, \mu_2)$$

and the variance-covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

where ρ is the correlation coefficient. The joint pdf is

$$f(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[(\frac{x_1-\mu_1}{\sigma_1})^2 - 2\rho(\frac{x_1-\mu_1}{\sigma_1})(\frac{x_2-\mu_2}{\sigma_2}) + (\frac{x_2-\mu_2}{\sigma_2})^2\right]\right)$$

for $x = (x_1, x_2) \in \mathbb{R}^2$

Bi-variate Normal distribution

- Each $X_i \sim norm(\mu_i, \sigma_i)$
- $X_1|X_2 = x_2 \sim norm\left(\mu_1 + \rho\sigma_1 \frac{x_2-\mu_2}{\sigma_2}, \sigma_1^2(1-\rho^2)\right)$
- $X_2|X_1 = x_1 \sim norm\left(\mu_2 + \rho\sigma_2 \frac{x_1-\mu_1}{\sigma_1}, \sigma_2^2(1-\rho^2)\right)$
- X_1 and X_2 are independent if and only if $\rho = 0$
- For any two constants a_1 and a_2,
  if $(X_1, X_2) \sim mvnorm(\mu, \Sigma)$ then
  $a_1X_1 + a_2X_2 \sim norm\left(a_1\mu_1 + a_2\mu_2, \sqrt{a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\rho\sigma_1\sigma_2}\right)$
- If $(X_1, X_2) \sim mvnorm(\mu, \Sigma)$ then $Z_1 = \dfrac{X_1 - \mu_1}{\sigma_1}$ and
  $Z_2 = \dfrac{1}{\sqrt{1-\rho^2}} \dfrac{X_2 - \mu_2}{\sigma_2} - \dfrac{\rho}{\sqrt{1-\rho^2}} \dfrac{X_1 - \mu_1}{\sigma_1}$ are two independent
  $norm(0, 1)$ random variables

Bi-variate Normal distribution

$X_i \sim norm(\mu_i, \sigma_i)$ for $i = 1, 2$ doesn't imply that the joint distribution is a bi-variate normal distribution.
Counter example: Let $X \sim norm(0, 1)$ and let

$$Y = \begin{cases} X & \text{if } X > 0, \\ -X & \text{if } X < 0 \end{cases}$$

then Y has a $norm(0, 1)$ distribution. But

$$X + Y = \begin{cases} 2X & \text{if } X > 0, \\ 0 & \text{if } X < 0 \end{cases}$$

which is not normally distributed in contradiction to the now-a-fact: the sum of two jointly normal r.v.'s is a normal r.v.

Limit Theorems

Weak and Strong Law of Large Numbers

If $X_1, X_2, \ldots$ are i.i.d.r.v.'s such that $\mathbf{E}|X_1| < \infty$ and $\mathbf{E}(X_1) = \mu$, and $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ then

- Weak Law of Large Numbers (WLLN)
  for every $\epsilon > 0$,
  $$\lim_{n\to\infty} P(\left|\bar{X}_n - \mu\right| < \epsilon) = 1$$
  Then we state that by saying $\bar{X}_n \to \mu$ in probability.
- Strong Law of Large Numbers (SLLN)
  for every $\epsilon > 0$,
  $$P(\lim_{n\to\infty}\{\omega \in S : \left|\bar{X}_n(\omega) - \mu\right| < \epsilon\}) = 1$$
  Then we state that by saying $\bar{X}_n \to \mu$ almost surely (with probability one).

Central Limit Theorem

If $X_1, X_2, \ldots$ are i.i.d.r.v.'s such that $\mathbf{E}(X_1) = \mu$ and $\mathbf{V}(X_1) = \sigma^2 < \infty$, and $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $Z_n = \dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ then

$$Z_n \to Z \text{ in distribution}$$

where $Z \sim norm(0, 1)$
That is,

$$\lim_{n\to\infty} F_{Z_n}(t) = \Phi(t) := \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2}} dx$$

Basics of Statistics

Point Estimation

If you model $X \sim dist(\theta_1, \theta_2, \ldots, \theta_k)$. To fully determine the model you sample (record) n independent instances of X so as to use them to estimate the parameters $\theta_1, \theta_2, \ldots, \theta_k$ which we denote $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$. You then use one of the following

- Method of moments
- Maximum likelihood method
- Expectation/Maximization method
- Bayesian method

Method of moments

If you model $X \sim dist(\theta_1, \theta_2, \ldots, \theta_k)$ and selected a simple random sample $x_1, x_2, \ldots, x_n$
Set

$$E(X^r) = \frac{1}{n}\sum_{i=1}^{n} x_i^r \text{ for } r = 1, 2, \ldots, k$$

and solve those k equations for $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$.

Maximum likelihood method

- The likelihood (probability) $L(x|\theta)$ or $L(\theta)$ of observing that simple random sample $x_1, x_2, \ldots, x_n$ of i.i.d. measurements is

$$L(\theta) := L(x_1, x_2, \ldots, x_n|\theta_1, \theta_2, \ldots, \theta_k) = \prod_{i=1}^{n} f(x_i|\theta_1, \theta_2, \ldots, \theta_k)$$

  by independence
- The maximum likelihood principle (due to Fisher) finds the MLE $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ that maximize the likelihood $L(\theta)$ of observing those observations $x_1, x_2, \ldots, x_n$.
- Some times we prefer to maximize $\ell(\theta) := \log(L(\theta)) = \sum_{i=1}^{n} \log(f(x_i|\theta))$ or minimize $-\ell(\theta)$
- Invariance property: If $\hat{\theta}$ is an MLE of θ then $g(\hat{\theta})$ is an MLE of $g(\theta)$

Bayesian method

The parameter θ (while has a specific but unknown value) is modeled as a random variable Θ due to the uncertainty about its value.* One can use a prior belief about it θ to assign wights $f_\Theta(\theta)$ to its possible values which might be just the uniform probability distribution if their is no specific belief about it values and so all the values are dealt with as being equally likely.

$$f_{\theta|x_1,\ldots,x_n}(\theta) = \frac{L(\theta|x_1,\ldots,x_n)f_\theta(\theta)}{\int_\Theta L(\theta|x_1,\ldots,x_n)f_\theta(\theta)d\theta}$$

or simply

$$\textit{posterior} \propto \textit{likelihood} \times \textit{prior}$$

and so

$$\mathbf{E}(h(\theta)) = \int_\Theta h(\theta)f_{\theta|x_1,\ldots,x_n}(\theta)d\theta$$

Note: we don't need to know the constant that makes L a joint pdf as it will cancel with itself from the denominator

Example

German Tank Problem

A number of Panther tanks were hunted down during WWII and their serial numbers were recorded, say, 86,43,19,183,128,252.
The task is to know the size of the German tank production.
For more info. visit
https://en.wikipedia.org/wiki/German_tank_problem

Example

German Tank Problem

The following is an excerpt from
https://en.wikipedia.org/wiki/German_tank_problem

Specific data [edit]

According to conventional Allied intelligence estimates, the Germans were producing around 1,400 tanks a month between June 1940 and September 1942. Applying the formula below to the serial numbers of captured tanks, the number was calculated to be 256 a month. After the war, captured German production figures from the ministry of Albert Speer showed the actual number to be 255.[3]

Estimates for some specific months are given as:[7]

Month	Statistical estimate	Intelligence estimate	German records
June 1940	169	1,000	122
June 1941	244	1,550	271
August 1942	327	1,550	342

Example

German Tank Problem

Let serial numbers, generally, be $x_1, x_2, \ldots, x_k$ were randomly sampled from the population of production. They can be ordered to $x_{(1)}, x_{(2)}, \ldots, x_{(k)}$ and let the total size of production be N which is a parameter. Thus,

$$P(X_{(k)} = m|N, k) = \frac{C_{k-1}^{m-1}}{C_k^N} \text{ for } m = k, \ldots, N$$

with

$$E(X_{(k)}) = \frac{k}{k+1}(N+1)$$

Example

German Tank Problem

By the method of moments, set $E(X_{(k)}) = x_{(k)}$ where the average of the observed maximum values is $x_{(k)}$ itself since it is observed once, which gives

$$\hat{N} = x_{(k)}\frac{k+1}{k} - 1$$

which is equal to $252 \times \frac{7}{6} - 1 = 293$ tanks in the example.

End of Set 2