Ch2. Random Variables

1. Random Variables
2. Discrete Random Variables
3. Continuous Random Variables
4. Expectation of a Random Variable
5. Jointly Distributed Random Variables
6. Moment Generating Functions
7. Limit Theorems

Random Variables

- (S, P) : Probability space from an experiment $\mathcal{E}$.
  However, it is convenient to assign number for each elementary outcome. Why? convenient to write, can count, do some arithematic operations, etc.

$$(S, P) \to (\mathbb{R}, P_X)$$

$$X : \omega \in S \to X(\omega) \in \mathbb{R}$$

Definition

A random variable (r.v) X is a function assigns a value $X(\omega) \in \mathbb{R}$ to each outcome $\omega \in S$.

- Hence, we can fully characterize a r.v by describing
  - $i)$ What values it takes and
  - $ii)$ Associated probability

Random Variables

- How to know probability for a r.v X?
  See inverse image of X and take probability on (S, P).
- Example: Coin tossing with $S = \{H, T\}$ and $P(H) = p$. Now define random variable X by assigning number 1 if head is observed, and 0 otherwise. Then,

Types of random variables

- X is a discrete r.v if it can take values on at most countable number of possible values.
- X is a continuous r.v. if the set of possible value is uncountable
- Countable / uncountable? If cardinality corresponds to integer points, then it is countable. For example, rational numbers ($x = m/n$ with gcd(m,n) $= 1$, $n \neq 0$) are countable.

- Simply speaking if X can take any value in an interval, then it is continuous.

Distribution function

- If X is uncountable, we cannot provide probability on each possible value of X can take.
- In general, cumulative distribution function(cdf) $F(\cdot)$ of the random variable X

$$F(b) := P_X(X \leq b) = P(\omega : X(\omega) \leq b)$$

characterizes probability of X. P_X is induced probability (measure) by random variable X.
(e.g)

$$F(b) = \begin{cases} 0 & \text{if } b < 0 \\ 1 - p & \text{if } 0 \leq b < 1 \\ 1 & \text{if } 1 \leq b \end{cases}$$

Properties of cdf

1. F is non-decreasing.

2. $\lim_{b \to -\infty} F(b) = 0, \lim_{b \to \infty} F(b) = 1$

3. F is right continuous in the sense that

$$\lim_{b_n \downarrow b} F(b_n) = F(b)$$

Properties of cdf

Remark

Conversely, any function φ defined on the real line satisfying above three properties, there is a random variable X with cdf φ (simple take $S = \mathbb{R}, X(\omega) = \omega$)

Remark

CDF is useful in probability calculation.

i) $P(a < X \leq b) = F(b) - F(a)$

ii) $P(X > b) = 1 - F(b)$

iii) *(left limit)*

$$P(X < b) = \lim_{n\to\infty} P\left(X \leq b - \frac{1}{n}\right) = \lim_{n\to\infty} F\left(b - \frac{1}{n}\right) = F(b-)$$

iv) *(jump size)*

$$P(X = b) = F(b) - F(b-)$$

Example: CDF

For the distribution function given by

$$F(x) = \begin{cases} 0, & x < 0 \\ x/2, & 0 \leq x < 1 \\ 2/3, & 1 \leq x < 2 \\ 11/12, & 2 \leq x < 3 \\ 1, & 3 \leq x, \end{cases}$$

Find

- $P(X < 3)$
- $P(X = 1)$
- $P(X > 1/2)$
- $P(2 < X \leq 4)$

Expectation of a R.V

Definition

$$\begin{cases} E(g(X)) = \sum_{x:p(x)>0} g(x)p(x) & discrete \\ E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx & continuous \end{cases}$$

It can be written in one formula using Lebesgue-Stieltjes integration

$$E(g(x)) = \int_{-\infty}^{\infty} g(x)dF(x)$$

Also, it is equivalent to write as

$$E(g(X)) = \int_{\omega \in S} g(X(\omega))dP(\omega)$$

(e.g, $E(g(X)) = \sum_{\omega \in S} g(X(\omega))P(\omega)$)

Expectation of a R.V

Indeed:

$$
\begin{aligned}
EX &= \sum_i x_i P_X(x_i) = \sum_i x_i P_X(X = x_i) \\
&= \sum_i x_i P(\{\omega \in S | X(\omega) = x_i\}) =: \sum_i x_i P(\omega \in S_i) \\
&= \sum_i x_i \sum_{\omega \in S_i} P(\omega) = \sum_i \sum_{\omega \in S_i} x_i P(\omega) \\
&= \sum_i \sum_{\omega \in S_i} X(\omega) P(\omega) = \sum_{\omega \in S} X(\omega) P(\omega)
\end{aligned}
$$

Example: Flip a coin twice and X denote the number of heads.
$E(X) = \sum_i x_i P_X(X = x_i) =$
$E(X) = \sum_\omega X(\omega) P(\omega) =$

Useful expectation formula for non-negative r.v

Theorem

For a non-negative r.v X

$$EX = \int_0^\infty P(X > t)dt$$

Graphically, it is

Useful expectation formula for non-negative r.v

Proof: 1. discrete case

$$EX = \sum_i x_i P(X = x_i) = \sum_i \int_0^{x_i} dt P(X = x_i)$$

$$= \sum_i \int_0^\infty 1_{\{t < x_i\}} \, dt \, P(X = x_i)$$

$$= \int_0^\infty \sum_i 1_{\{t < x_i\}} P(X = x_i) dt = \int_0^\infty P(X > t) dt$$

2. Continuous case where the density function is $f(x)$.

$$\int_0^\infty P(X > t) dt = \int_0^\infty \int_t^\infty f(x) dx dt$$

$$= \int_0^\infty \int_0^\infty f(x) 1_{\{x \geq t\}} dx dt = \int_0^\infty \left\{ \int_0^\infty f(x) 1_{\{x \geq t\}} \, dt \right\} dx$$

$$= \int_0^\infty f(x) t|_0^x dx = \int_0^\infty f(x) x dx = EX$$

Extension

Remark

If X taking values in $\mathbb{N} = \{0, 1, \cdots\}$

$$EX = \sum_{n=0}^{\infty} P(X > n)$$

Remark

Since random variable X *can be written as*

$$X = X^+ - X^-$$

$$X^+ = X1_{\{X \geq 0\}}, \quad X^- = -X1_{\{X<0\}},$$

for "any" real-valued random variable X

$$EX = \int_0^{\infty} P(X > t)dt - \int_{-\infty}^{0} P(X \leq t)dt$$

Random vectors and joint distributions

- Consider we perform n experiments at the same time. Then, the sample space is given by Cartesian product

$$\begin{aligned} S &= S_1 \times \cdots \times S_n \\ &= \{\boldsymbol{\omega} := (\omega_1, \cdots, \omega_n) | \omega_1 \in S_1, \cdots, \omega_n \in S_n\} \end{aligned}$$

- We can still define probability model (S, P) by using three axioms where $\boldsymbol{\omega}$ is treated as simple outcome.

Definition

An n-dim'l random vector is a function

$$\mathbf{X} = (X_1, \cdots, X_n) : S_1 \times \cdots \times S_n \to \mathbb{R}^n$$

such that each component $X_i : S_i \to \mathbb{R}$ is a r.v.

Random vectors and joint distributions

1. Distribution of $\mathbf{X}$ is given by

$$F_{\mathbf{X}}(x_1, \cdots, x_n) = P_X(X_1 \leq x_1, \cdots, X_n \leq x_n)$$

$$= P(X_1^{-1}(-\infty, x_1] \times X_2^{-1}(-\infty, x_2] \times \cdots \times X_n^{-1}(-\infty, x_n])$$

2. For discrete r.v pmf is given by

$$P(X_1 = x_1, \cdots X_n = x_n)$$

3. For continuous r.v joint pdf is given by

$$P(X_1 \in A_1, \cdots, X_n \in A_n)$$

$$= \int_{A_n} \int_{A_{n-1}} \cdots \int_{A_1} f(x_1, \cdots, x_n) dx_1 dx_2 \cdots dx_n$$

4. Remark that pdf always satisfy

$i)$non-negative

$ii)$add up to 1

Independent random variables

- Recall that in (S, P), two events C and D are independent iff

$$P(C \cap D) = P(C)P(D)$$

Definition

X and Y are independent if

$$P_{X,Y}(X \in A, Y \in B) = P_X(X \in A)P_Y(Y \in B)$$

for any subsets A, B in $\mathbb{R}$. This is equivalent to say in (S, P)

$$P(X^{-1}(A) \cap Y^{-1}(B)) = P(X^{-1}(A))P(Y^{-1}(B)).$$

Independent random variables

- Observe that any subset in $\mathbb{R}$ can be as the countable union/intersection/complement of the form $(-\infty, x]$. For example,

  $$(a, b] = (-\infty, b] - (-\infty, a]$$

  $$b = \bigcap_{n=1}^{\infty} \left(b - \tfrac{1}{n}, b\right]$$

  $$(-\infty, b) = \cup_{n=1}^{\infty} \left(-\infty, b - \tfrac{1}{n}\right]$$

  In this sense, we say that interval of the form $(-\infty, x]$ generates any subset in $\mathbb{R}$.
- Independence of two random variables X and Y equals to

  $$P_{X,Y}(X \in (-\infty, x], Y \in (-\infty, y])$$

  $$= P_X(X \in (-\infty, x]) P_Y(Y \in (-\infty, y])$$

  $$\boxed{F_{X,Y}(x, y) = F_X(x) F_Y(y)}$$

Useful properties

If X and Y are independent, then

1. (CDF) $F_{X,Y}(x,y) = F_X(x)F_Y(y)$
2. (PDF when exists)

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

3. (MGF)

$$M_{X,Y}(t_1,t_2) = M_X(t_1)M_Y(t_2)$$

4. $E(g(X)h(Y)) = E(g(X))E(h(Y))$
5. $\mathrm{Cov}(X,Y) = 0$ (but not conversely)

Before introducing MGF, let us first see how to understand the sum of random variables.

Sum of random variables

For random variables $X_1, \dots, X_n$ defined on the same probability space $(S, \mathcal{F}, P)$,

$$Z = X_1 + \cdots + X_n$$

is undersood as

$$Z(\omega) = X_1(\omega) + \cdots + X_n(\omega)$$

1. $E(X_1(\omega) + \cdots + X_n(\omega)) = \sum_\omega (X_1(\omega) + \cdots + X_n(\omega))P(\omega)$
$$= EX_1 + \cdots + EX_n$$
2. $\mathrm{Var}(X_1 + \cdots X_n) = \mathrm{Cov}(X_1 + \cdots + X_n, X_1 + \cdots + X_n)$
$$= \sum_i \sum_j \mathrm{Cov}(X_i, X_j) = \sum_i \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, Y_j)$$
3. If $X_i's$ are independent, then
$$\mathrm{Var}(X_1 + \cdots + X_n) = \sum_i \mathrm{Var}(X_i)$$

Moment generating function (MGF)

$$\boxed{\phi(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} dF(x)}$$

- It is called the moment generating function b/c it carries information about all moments.

$$\phi'(t) = \frac{d}{dt}E(e^{tX}) = E\left(\frac{d}{dt}e^{tX}\right) = E(Xe^{tX}) \Rightarrow \phi'(0) = EX$$

$$\boxed{EX^n = \phi^{(n)}(0)}$$

- (Uniqueness of mgf) MGF uniquely determines the (cumulative) distribution of random variable.

$$M_X(t) = M_Y(t), \quad \forall t \iff X \stackrel{d}{=} Y$$

Moment generating function (MGF)

- If X and Y are independent, then

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

- If $X_1, \cdots X_n$ are independent, then

$$M_{X_1+\cdots+X_n}(t) = M_1(t)M_2(t)\cdots M_n(t)$$

- $E(e^{-tX}) = \phi(-t)$ is called Laplace transformation.
- However, it is possible to have all finite moments but MGF does not exist. (example : log-normal distribution)
- MGF may not exist, e.g Cauchy distribution.
- For multivariate case $\mathbf{X} = (X_1, \ldots, X_n)$, $\mathbf{t} = (t_1, \ldots, t_n)$,

$$M_{\mathbf{X}}(\mathbf{t}) := E(e^{\mathbf{t}'\mathbf{X}}) = E(e^{t_1X_1+\ldots+t_nX_n})$$

Popular Discrete Random Variables

- Bernoulli, Binomial, Geometric, Negative Binomial and Poisson distribution will be reviewed.
- Focus is to understand above distributions via Bernoulli process.

1. Bernoulli distribution

- Experiment: (Bernoulli trial) only two possible outcomes; either S or F.
- $S = \{\mathrm{S}, \mathrm{F}\}$ with $P(\mathrm{S}) = p$ (success probability)
- Bernoulli random variable X is given by

$$X = \begin{cases} 1 & \text{if success with } p \\ 0 & \text{if failure with } 1-p \end{cases}$$

- pmf : $f(x) = p^x(1-p)^{1-x}, \quad x = 0, 1$
- mgf : $M_X(t) = pe^t + q, \quad -\infty < t < \infty$
- $EX = 1 \cdot p + 0 \cdot (1-p) = p$
- $\mathrm{Var}(X) = pq$
- Notation : $X \sim$ Bernoulli(p)

Bernoulli process/trials

- Bernoulli process/ trials : Perform Bernoulli trial independently and identically many times. Hence we will have sequence of observations

$$\{X_1, X_2, \cdots\} \sim \text{IID Bernoulli(p)}$$

- Examples:
  -

2. Binomial distribution

- Experiment: suppose we have observed n independent Bernoulli trials. Then,

$$S = \{(\mathrm{S}, \mathrm{S}, \mathrm{S}, \ldots, \mathrm{S}), (\mathrm{S}, \mathrm{S}, \mathrm{S}, \ldots, \mathrm{F}), \ldots, (\mathrm{F}, \mathrm{F}, \ldots, \mathrm{F})\}$$

- Binomial random variable

$$X\text{: record the number of success}$$

- pmf is given by

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, 1, 2, \ldots, n$$

- Representational definition

$$\boxed{X \stackrel{d}{=} X_1 + \cdots + X_n}$$

  where X_i if IID Bernoulli(p) random variables.

- Notation : $X \sim$ Bin(n, p)

2. Binomial distribution

- $EX = E(X_1 + \ldots + X_n) = np$
- $\mathrm{Var}(X_1 + \ldots + X_n) = np(1-p)$
- $M_X(t) = E(e^{t(X_1+\cdots+X_n)}) = E(e^{tX_1}e^{tX_2}\cdots e^{tX_n})$
  $= E(e^{tX_1})\cdots E(e^{tX_n}) = \{E(e^{tX_1})\}^n = (pe^t + q)^n$
- Example: Diskettes produced by some company defective with probability .01. Packages of 10 diskettes are sold. Money-back guarantee if at most 1 defective (i.e $>= 2$ defective, then refund). Suppose 3 packages are purchased. Find the probability that exactly 1 will be refunded.

3. Geometric distribution

- Experiment: observe Bernoulli trials (countably many times)
- Geometric random variable

  $$X \,:\, \text{Total } \# \text{ of trials until the first success}$$

- If the simple outcome is given by $\mathrm{F}, \mathrm{F}, \mathrm{F}, \mathrm{F}, \mathrm{S}$, then $X = 5$.
- pmf : $P(X = n) = (1-p)^{n-1}p, \quad n = 1, 2, \ldots$
  Indeed pmf:

- Notation : $X \sim \text{Geo}(p)$
- $EX = \sum_{n=1}^{\infty} nq^{n-1}p = \frac{1}{p}$
- $\mathrm{Var}(X) = \frac{1}{p^2} - \frac{1}{p} \quad (\because EX^2 = \frac{2}{p^2} - \frac{1}{p})$

3. Geometric distribution

- MGF is calculated as

$$E(e^{tX}) = \sum_{n=1}^{\infty} e^{tn} p q^{n-1} = \frac{p}{q} \sum_{n=1}^{\infty} (e^t q)^n = \frac{p}{q} \frac{e^t q}{1 - q e^t} = \frac{p e^t}{1 - q e^t}$$

  if $|e^t q| < 1$. Therefore, MGF exists for $t < -\log q$.

- Depending on the context, instead of total number of trials, Geometric distribution is defined as the number of failures required to observe the success. In this case

$$Y = X - 1$$

$$P(Y = k) = P(X = k + 1) = q^k p, \quad k = 0, 1, \ldots.$$

4. Negative Binomial distribution

- Experiment: observe Bernoulli trials (countably many times)
- Negative Binomial random variable

$$X : \text{Total \# of trials required till } r^{th} \text{ success}$$

- If $r = 3$ and simple outcome is $\mathrm{F, S, F, S, F, F, F, S, F, F, S}$, then $X =$
- pmf is given by

$$\begin{aligned} P(X = n) &= P((r-1)\text{success out of } (n-1) \text{ trials}) \\ &= \binom{n-1}{r-1} p^{r-1} q^{(n-1)-(r-1)} \cdot p \\ &= \binom{n-1}{r-1} p^r q^{n-r}, \quad n = r, r+1, \ldots \end{aligned}$$

- Relation to Geo(p)

$$X = Y_1 + Y_2 + \cdots + Y_r, \quad Y_i \sim \mathrm{IID} \;\; \mathrm{Geo}(p)$$

4. Negative Binomial distribution

- $EX = E(Y_1 + \cdots + Y_r) =$
- $\mathrm{Var}(X) = r \cdot \mathrm{Var}(Y_1) =$
- mgf is calculated as

$$M_X(t) = \{E(e^{tY_1})\}^r = \frac{(pe^t)^r}{(1 - e^t q)^r}$$

  provided if $e^t q < 1$.
- Notation : Negbin(r,p)
- We can also define Negative binomial random variable as the number of **failures** required to observe rth success.

$$Y = X - r, \quad X \sim Negbin(r, p)$$

$$\begin{aligned} P(Y = k) &= P(X = k + r) \\ &= \binom{k + r - 1}{r - 1} p^r q^{r+k-r} \\ &= \binom{k + r - 1}{k} p^r q^k, \quad k = 0, 1, 2 \cdots \end{aligned}$$

4. Negbin Example: Banach match problem

At all times, a pipe-smoking mathematician carries 2 matchboxes - 1 in his left-handed pocket and 1 in his right-hand pocket. Each time he needs a match, he is equally likely to take it from either side. It is assumed that both matchboxes initially contained N matches. When one matchbox is discovered empty, find the probability that the other one has exactly k matches for $k = 0, 1, 2, \ldots, N$.

Negative binomial expansion

- Negative binomial distribution is named after negative binomial expansion. Recall Binomial theorem:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

- Negative binomial expansion theorem. From Taylor expansion

$$(x+a)^{-r} = \sum_{k=0}^{\infty} (-1)^k \binom{k+r-1}{k} x^k a^{-r-k}, \quad a < 1$$

It further equals to

$$\sum_{k=0}^{\infty} \binom{-r}{k} x^k a^{-r-k}$$

like Binomial expansion.

Negative binomial expansion

- Indeed:

$$\binom{k+r-1}{k} = \frac{(k+r-1)!}{k!(r-1)!} = \frac{(k+r-1)(k+r-2)\dots r}{k!}$$

$$= (-1)^k \frac{(-r)(-r-1)\dots(-r-k+1)}{k!} = (-1)^k \binom{-r}{k}$$

- Therefore, Negative binomial distribution defined as the number of failures till observe rth success agree with negative binomial coefficient up to constant $(-1)^k$.

Keep in mind

We can define popular random variables through Bernoulli process. Suppose that $\{X_i\}$ are IID Bernoulli(p) random variables, then

1. Total # of success: Bin(n, p)

$$X = X_1 + \cdots + X_n \sim \text{Bin(n,p)}$$

2. Waiting time til r^{th} success = Total # of trials till r^{th} success.

$$W_r = \min\{n : X_1 + \cdots + X_n \geq r\}$$

$$W_r \sim Negbin(r, p)$$

3. Inter-success (Inter-arrival) time in Bernoulli process.

$$W_1, W_2 - W_1, W_3 - W_2, \ldots, \sim \text{ IID Geo}(p)$$

Counting process

- Later on, we will define counting process. Most important example of counting process is Poisson process.
- In counting process, we are interested in both total number of success and waiting time. Graphically
- We will take limit $n \to \infty$ in the Bernoulli process to introduce Poisson, Exponential and Gamma distributions.

5. Poisson distribution

- Poisson distribution can be understood as the approximation of Binomial distribution with small success probability p, but moderate np (it happens indeed). For example
  - # of misprints on a page of a book.
  - # of people in a community who survive to age 100.
  - # of customers entering a post office on a given day
- Bin(n, p) with $n \to \infty$ for small p but moderate np, that is,

$$\boxed{\lim_{n\to\infty} np = \lambda > 0}$$

- This approximation is useful because calculating Binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!(n-k!)}$$

  is not easy at all for large n and k.
- Approximation to Poisson distribution

5. Poisson distribution

By using facts

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}, \quad e^x = 1 + x + \frac{x^2}{2!}\cdots = \sum_{n=0}^{\infty}\frac{x^n}{n!}$$

When $n \to \infty$, $np = \lambda$ gives

$$P(X = i) = \frac{n!}{i!(n-i!)}p^i(1-p)^{n-i}$$

5. Poisson distribution.

- Notation: $X_i \sim$ Poisson(λ), $\lambda > 0$
- pmf is given by

$$P(X = i) = \frac{e^{-\lambda}\lambda^i}{i!}, \quad i = 0, 1, 2, \ldots.$$

Indeed:

- $EX = \lambda$, $\mathrm{Var}(X) = \lambda$
- Intuitively $X \approx$ Bin(n, p) with $np = \lambda$, hence
  $EX \approx np = \lambda, \quad \mathrm{Var}(X) \approx np(1-p) = \lambda$
- MGF is given by

$$M_X(t) = \exp\{\lambda(e^t - 1)\}$$

Example: Poisson distribution

A person purchased 50 lottery tickets with winning probability of $1/100$. Find the probability that s/he has at least two wins.

Sol) Exact probability:

Approximation:

Poisson paradigm (Example 2.47)

Recall that

$$\text{Bin(n,p)} \approx \text{possion}(\lambda = np)$$

holds for IID Bernoulli process. Poisson paradigm states that this approximation holds even though IID assumptions are violated.

$i)$ indep trials with different success probability. That is, $X_i \sim \text{Bernoulli}(p_i)$ process. Then, as $n \to \infty$,

$$X_1 + \ldots + X_n \approx \text{Poisson}\left(\sum_{i=1}^{n} p_i\right)$$

Because, MGF of X_i is given by

$$p_i e^t + 1 - p_i = 1 + p_i(e^t - 1) \approx \exp(p_i(e^t - 1))$$

using $e^x \approx 1 + x$. Thus,

$$E(e^{t(X_1+\cdots+X_n)}) = \prod_{i=1}^{n} E(e^{tX_i}) \approx \prod_{i=1}^{n} \exp(p_i(e^t - 1))$$

$$= \exp\left(\sum_{i=1}^{n} p_i(e^t - 1)\right) = \text{ MGF of Poisson}\left(\sum_{i=1}^{n} p_i\right)$$

Poisson paradigm

$ii)$ Bernoulli process with dependent trials. Recall matching problem with event

$$A_i = i^{th} \text{ person get his own hat}$$

and denote $X_i = 1$ when A_i happens or 0 otherwise. Then, we have

$$P(A_i) = \frac{1}{n}, \quad P(A_i|A_j) = \frac{1}{n-1}, j \neq i,$$

so that X_i's are **dependent**. Using inclusion-exclusion formula

$$P(\text{no one get his own hat}) = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} - \cdots + \frac{(-1)^n}{n!} \to e^{-1},$$

as $n \to \infty$

- It can be understood as Possion approximation. $np = n \cdot \frac{1}{n} = 1$ gives $X_1 + \ldots + X_n \approx Possion(1)$, hence

$$\lim_{n \to \infty} P(X_1 + \ldots + X_n = 0) = e^{-1}.$$

6. Exponential distribution

- Recall that in the Bernoulli process, waiting time between success (inter-arrivals) is given by

$$W_1, W_2 - W_1, \cdots \sim \text{ IID Geo(p).}$$

- Continuous analogue is called the Exponential distribution. Exact derivation will be done in Poisson process.
- Notation: $Y \sim \text{Exp}(\lambda)$ with rate λ (or mean $1/\lambda$).
- pdf: $\lambda e^{-\lambda y} 1_{\{y \geq 0\}}$
- cdf: $F(y) = P(Y \leq y) = 1 - e^{-\lambda y}$
- MGF: $M_Y(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}, \quad t < \lambda$
- $E(X) = 1/\lambda$ and $\text{Var}(X) = 1/\lambda^2$.
- In fact $\lfloor Y \rfloor$ (greatest interger) is Geo(p) with $p = 1 - e^{-\lambda}$.

Waiting time till r^{th} success = Gamma(r,λ)

- Similarly, waiting time till r^{th} success is defined as

$$Y = Y_1 + Y_2 + \cdots + Y_r,$$

  and when $Y_i \sim \text{Exp}(\lambda)$ it is known that Y follows Gamma distribution with parameters $r > 0$ and $\lambda > 0$.

- It is related to Gamma function

$$\Gamma(\alpha) := \int_0^\infty y^{\alpha-1} e^{-y} dy, \quad \alpha > 0$$

  Basic properties include:

  $i)$ $\Gamma(1) = \int_0^\infty e^{-x} dx = 1$

  $ii)$ $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1), \alpha > 1$

  $iii)$ $\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)(n-2)\Gamma(n-2) = \cdots = (n-1)(n-2)\cdots\Gamma(1) = (n-1)!$

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

7. Gamma distribution

- Notation: $X \sim \text{Gamma}(r, \lambda)$, $r > 0$ and $\lambda > 0$
- pdf : $\dfrac{1}{\Gamma(r)}\lambda^r x^{r-1} e^{-\lambda x} 1_{\{x>0\}}$
- $EX = \dfrac{r}{\lambda}$
- $\mathrm{Var}(X) = \dfrac{r}{\lambda^2}$
- MGF: $M_X(t) = (1 - \frac{t}{\lambda})^{-r}, \quad t < \lambda$

Summary of Bernoulli process $\{X_i, i \geq 1\}$

1. Total # of success

$$X_1 + \cdots + X_n \sim \text{Bin}(n, p)$$

2. # trials till r^{th} success

$$W_r = \min\{n : X_1 + \cdots + X_n \geq r\} \sim \text{Negbin}(r, p)$$

3. # of trials till next success.

$$W_1, W_2 - W_1, W_3 - W_2, \ldots \sim \text{ IID Geo(p)}$$

When $n \to \infty$ with $np \approx \lambda > 0$ (continuous analogue)

4. Total # of success: Bin$(n, p) \approx$ Possion(λ)
5. Inter-arrival time: IID Geo(p) $\approx$ IID Exp(λ)
6. Waiting time till r^{th} arrival: Negbin$(r, p) \approx$ Gamma(r, λ)

We will study further about this approximation in Chapter 5.
Poisson process.

Basic Limit Theorems

- Law of Large Numbers(LLN)
- Central Limit Theorem (CLT)

Long-run relative frequency converges to probability defined through axioms of probability.

Law of large numbers

- We want to say rigorously when the sample size is fairly large,

$$\overline{X} \approx \mu := \int x dF(x),$$

that is, for random sample $X_1, \ldots, X_n$, we want to say

$$\lim_{n\to\infty} \frac{X_1 + \ldots + X_n}{n} = \mu$$

- However, above statement makes no sense because it is a random sample (depending on ω), so different realization may produce different numbers. It suggests that limit also need probabilistic argument.
- Two types of Law of large numbers

Weak Law of Large Numbers (WLLN)
Strong Law of Large Numbers (SLLN)

Law of Large Numbers

Definition (WLLN)

Let $X_1, \ldots, X_n$ *be a sequence of IID random variables with finite mean* $EX_i = \mu$*, then for any* $\epsilon > 0$

$$P\left\{ \left| \frac{X_1 + \cdots + X_n}{n} - \mu \right| \geq \epsilon \right\} \to 0 \quad \textit{as} \;\; n \to \infty$$

Definition (SLLN)

Strong law of large numbers

$$P\left(\frac{X_1 + \cdots + X_n}{n} \to \mu \;\; \textit{as} \;\; n \to \infty \right) = 1$$

LLN & Axioms of Probability

LLN is important because it proves that long-run relative frequency indeed converges to $P(A)$.

- Take

$$X_i = \begin{cases} 1 & \text{if event } A \text{ happen} \\ 0 & \text{o.w} \end{cases}$$

Then, by applying LLN we have

Central Limit Theorem

- LLN says sample average is close to population mean when sample size is very large.
- We want to say more precisely, how fast it converges to population mean.
- Central limit theorem says the speed of convergence (in terms of sample size).

Definition

Let $X_1, \cdots, X_n$ be iid random variables from a pop^n with mean μ and variance σ^2. Then,

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1),$$

where $\xrightarrow{d}$ represents convergence in distribution.

CLT

- Convergence in distribution means that

$$P(Z_n \leq a) \to P(N(0,1) \leq a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$$

for any $a \in \mathbb{R}$. Equivalently (due to Lévy),

$$M_{Z_n}(t) \to M_Z(t), \quad \forall t \in \mathbb{R}$$

- Remark that CLT hols for any distributions with finite second moment. IID assumptions can be relaxed.
- In terms of sample average, it reads as

$$\overline{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

CLT for IID Bernoulli(p)

- For example, consider $X_1, \ldots, X_n \sim$ IID Bernoulli(p),

$$\frac{X_1 \cdots + X_n - np}{\sqrt{npq}} \approx N(0,1)$$

Or, equivalently,

$$\frac{\overline{X} - p}{\sqrt{pq}/\sqrt{n}} \approx N(0,1) \qquad \frac{\text{sample average - mean}}{s.d/\sqrt{n}} \approx N(0,1)$$

- (Normal approximation to Binomial)
When $np \to \infty$,

$$X \sim \text{Bin}(n,p) \approx \mathcal{N}(np, npq).$$

Therefore, we can approximate Binomial probability as

$$P(X \leq a) \approx P\left(Z \leq \frac{a - np}{\sqrt{npq}}\right)$$