0 math voodoo $e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + \dots$ R=∞

$\int u\,dv = uv - \int v\,du$, partial fractions, $f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n, |x-a| < (\lim_{n\to\infty} |\frac{c_{n+1}}{c_n}|)^{-1}$, $\sqrt{a^2 \pm (x+b)^2} \Rightarrow x+b = a\sin\theta$ tan sec

1 Combinatorial Analysis $\int_a^b \int_c^d f(x,y)\,dy\,dx \Leftrightarrow f(x,y) = g(x)h(y)$

Generalised Principle of Counting $= \int_a^b g(x)dx \int_c^d h(y)dy$

For r experiments, where the i-th experiment can result in any of n_i possible outcomes, we have $n_1 \cdot n_2 \cdot \ldots \cdot n_r$ outcomes.

Permutations

- $n!$ arrangements of n distinct objects
- $\frac{n!}{n_1!n_2!\ldots n_r!}$ arrangements of n objects of which n_1 are alike, n_2 are alike, ..., n_r are alike

Combinations

- $\binom{n}{r} = \frac{n!}{(n-r)!r!}$, for $r \le n$
- $\binom{n}{r}$ subsets of size r from a set of size n.
- $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}, 1 \le r \le n$
- $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$
- $\binom{n}{n_1, n_2, \ldots, n_r}$ divisions of n distinct objects into r *distinct* groups sized $n_1, n_2, \ldots, n_r$ where $n_1 + n_2 + \ldots + n_r = n$.
- For $n_1 + \ldots + n_r = n$ and $n_i \ge 0$, $(x_1 + \ldots + x_r)^n =$
$$\sum_{(n_1,\ldots,n_r):n_1+\ldots+n_r=n} \binom{n}{n_1,\ldots,n_r} x_1^{n_1} x_2^{n_2} \ldots x_n^{n_r}$$
- Integer solutions for $n_1 + n_2 + \ldots + n_r = n, n_i > 0$:
  - $\binom{n+r-1}{r-1}$ distinct vectors $(n_1, n_2, \ldots, n_r) \in \mathbb{N}^r$
  - $\binom{n-1}{r-1}$ distinct vectors $(n_1, n_2, \ldots, n_r) \in \mathbb{N}^{*r}$

2 Axioms of Probability

Event $E \subseteq$ Sample Space S

- Sample Space S Set of all possible outcomes of an experiment
- Mutually Exclusive Events $E, F \Leftrightarrow E \cap F = \phi$
- $E \subseteq F \iff \forall x \in E (x \in F)$
  $\iff E \cap F = E$
- $E = F \Leftrightarrow E \subseteq F \wedge F \subseteq E$
- Properties of $\cup$ and $\cap$:
  Commutative. Associative. Distributive over the other.
- DeMorgan's Laws
$$(\bigcup_{i=1}^{n} E_i)^c = \bigcap_{i=1}^{n} E_i^c \text{ and } (\bigcap_{i=1}^{n} E_i)^c = \bigcup_{i=1}^{n} E_i^c$$

Probability: Limiting Relative Frequency

For an experiment performed repeatedly many times under the same conditions, the probability of an event E happening is defined:
$$P(E) = \lim_{n\to\infty} \frac{n(E)}{n}$$
where $n(E)$ = frequency of E in n repetitions of the experiment

Kolmogorov's Axioms of Probability

For any event E, assume $P(E)$ is defined and satisfies axioms:

1. $0 \le P(E) \le 1$
2. $P(S) = 1$
3. For any sequence of mutually exclusive events $E_1, E_2, \ldots$,
$$P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$$

$P(\cdot)$ satisfying the three axioms is known as a probability function.

Law of Large Numbers $\frac{1}{1-x} = 1 + x + x^2 + \dots + x^n + \dots$ R=1

For $X_1, X_2, \ldots$ taken independently from a population, $\sum_{i=1}^{n} \frac{X_i}{n}$ almost surely (ie. with probability 1) converges to $E(X_i)$.

Probabilities!!!

1. $P(E^c) = 1 - P(E)$
2. $E \subseteq F \Rightarrow P(E) \le P(F)$
3. $\bigcup_{i=1}^{n} E_i = E_1 \cup E_1^c E_2 \cup E_1^c E_2^c E_3 \cup \ldots \cup E_1^c \ldots E_{n-1}^c E_n$
4. Principle of Inclusion Exclusion (PIE)
$$P(\bigcup_{i=1}^{n} E_i) = P(E_1 \cup E_2 \cup \ldots \cup E_n)$$
$$= \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} E_{i_2}) + \ldots + (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r} P(E_{i_1} E_{i_2} \ldots E_{i_r}) + \cdots + (-1)^{n+1} P(E_1 E_2 \ldots E_n)$$
5. $$P(\bigcup_{i=1}^{n}) E_i \le \sum_{i=1}^{n} P(E_i)$$
$$\ge \sum_{i=1}^{n} P(E_i) - \sum_{j<i} P(E_i E_j)$$
$$\le \sum_{i=1}^{n} P(E_i) - \sum_{j<i} P(E_i E_j) + \sum_{k<j<i} P(E_i E_j E_k)$$

Sample Space with Equally Likely Outcomes

$S = \{e_1, e_2, \ldots, e_n\}$ where $P(\{e_1\}) = \cdots = P(\{e_n\})$

$P(\cdot)$ satisfies axioms $\Rightarrow \forall E \subseteq S, P(E) = \frac{|E|}{|S|}$

Probability as a Continuous Set Function

- $E_1, E_2, \ldots$ is an increasing sequence if $E_1 \subset E_2 \subset \ldots$
- Increasing Sequence $\{E_n, n \ge 1\} \Rightarrow$
$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n) \equiv P(\bigcup_{i=1}^{\infty} E_i)$$
- $E_1, E_2, \ldots$ is a decreasing sequence if $E_1 \supset E_2 \supset \ldots$
- Decreasing Sequence $\{E_n, n \ge 1\} \Rightarrow$
$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n) \equiv P(\bigcap_{i=1}^{\infty} E_i)$$

Probability as a Measure of Belief

Definition of probability as a measures of belief should still satisfy all of Kolmogorov's three axioms.

3 Conditional Probability and Independence

Conditional Probability

Probability E occurs given that F occurred is denoted $P(E|F)$
$$P(E|F) = \frac{P(EF)}{(F)}, P(F) > 0$$

Conditional probability is a probability because it satisfies Kolmogorov's three axioms.

- $0 \le P(E|F) \le 1$
- $P(S|F) = 1$
- If E_i for $i = 1, 2, \ldots$ are mutually exclusive events, then
$$P(\cup_{i=1}^{\infty} E_i | F) = \sum_{i=1}^{\infty} P(E_i|F)$$

Given that F occurred, occurrence of E is in EF. Only need to focus on F, the *reduced sample space*.

- $P(EF) = P(F)P(E|F)$
- $P(E_1 E_2 \ldots E_n) = P(E_1)P(E_2|E_1)\ldots P(E_n|E_1 \ldots E_{n-1})$
- For $F_1, F_2, \ldots, F_n$ partitioning S,
$$P(E) = \sum_{i=1}^{n} P(EF_i) = \sum_{i=1}^{n} P(F_i)P(E|F_i)$$
- Baye's Theorem: E occurred. Which F_j most likely led to E:
$$P(F_j|E) = \frac{P(EF_j)}{P(E)} = \frac{P(F_j)P(E|F_j)}{P(E)} = \frac{P(F_j)P(E|F_j)}{\sum_{i=1}^{n} P(F_i)P(E|F_i)}$$
- Odds of $E = \frac{P(F)}{P(F^C)}$

Independent Events

- E is independent of $F \iff F$ is independent of E
  $\iff P(E|F) = P(E)$
  $\iff P(EF) = P(E)P(F)$
  $\iff P(E|F) = P(E)$
  $\iff E$ and F^c are independent
- n events $E_1, E_2, \ldots, E_n$ are independent $\iff$ Any subset of the events are independent

4 Random Variables

Random variable maps sample space to a countable number of possible values.
$$X : S \mapsto \mathbb{R}$$

For some true value x_0 being measured by X, **mean squared error** $MSE = E[(X - x_0)^2] = Var(X - x_0) + [E(X - x_0)]^2 = Var(X) + (bias)^2$

- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
- $$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + \sum_{i,j \in \{1..n\}, i \ne j} Cov(X_i, X_j)$$
- For independent X_i,
$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

4.1 Discrete Random Variables

PMF of a discrete random variable is $p(a) = P(X = a)$. $p(a) > 0$ for at most a countable number of values of a, and $\sum_{i=1}^{\infty} p(x_i) = 1$

CDF of a discrete random variable is $F(a) = P(X \le a) = \sum_{\text{all} x \le a} p(x)$ $F(a)$ is defined $\forall a \in \mathbb{R}$, and is a step function at every value of a for which $p(a) > 0$

$E(X)$ is the average of all possible values of X, weighted by the probabilities. $E(X) = \sum_{x} xp(x)$.

For nonnegative integer-valued X, $E(X) = \sum_{i=1}^{\infty} P(X \ge i)$.

For real-valued function $g(x)$, $E[g(X)] = \sum_{i} g(x_i)p(x_i)$

$E(aX + b) = aE(X) + b$, for some $a, b \in \mathbb{K}$

n-th Moment of X: $E(X^n) = \sum_{x} x^n p(x)$

For random variable X with mean $\mu = E(X)$,
$$Var(X) = E[(X - \mu)^2] = \sum_{x} (x - \mu)^2 p(x) = E(X^2) - [E(X)]^2$$

$Var(aX + b) = a^2 Var(X)$, for some $a, b \in \mathbb{K}$

Bernoulli X has probability of success p if
$$P(X = x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \end{cases}$$
$X \sim Bernoulli(p), E(X) = p, Var(X) = p(1 - p)$

Binomial X is the number of successes in n independent Bernoulli trials, each trial resulting in success with probability p.
$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \forall k \ge 0$$
$$= \frac{(n - k + 1)p}{k(1 - p)} P(X = k - 1), \forall k \ge 1$$
$X \sim Bin(n, p), E(X) = np, Var(X) = np(1 - p)$

$X \sim Bin(n, p), Y \sim Bin(m, p) \Rightarrow X + Y \sim Bin(n + m, p)$

Geometric X is the number of independent Bernoulli trials performed until a success is obtained (inclusive of the success), each trial resulting in success with probability p.
$$P(X = k) = (1 - p)^{k-1} p, \forall k \ge 1$$
$X \sim Geometric(p), E(X) = \frac{1}{p}, Var(X) = \frac{1-p}{p^2}$

Negative Binomial X is the number of independent Bernoulli trials performed until m successes are obtained (inclusive of the last success), each trial resulting in success with probability p. For $k \ge m$,
$$P(X = k) = \frac{(k - 1)!}{(m - 1)!(k - m)!} (1 - p)^{k-m} p^m$$
$X \sim NB(m, p), E(X) = n/p, Var(X) = \frac{m(1-p)}{p^2}$

Hypergeometric X is the number of white balls selected in a sample of n balls without replacement from an urn of N balls of which m are white and $N - m$ are black. For $0 \le i \le n \wedge n - (N - m) \le i \le \min(n, m)$,
$$P(X = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$$
$X \sim Hypergeometric(N, m, n), E(X) = \frac{nm}{N}$, $Var(X) = np(1 - p)(1 - \frac{n-1}{N-1})$ where $p = \frac{m}{N}$

$n << N \Rightarrow X \sim Bin(n, \frac{m}{N}), Var(X) \approx np(1 - p)$

Poisson X has parameter λ if for some $\lambda > 0$, for $k \ge 0$,
$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$
$X \sim Poisson(\lambda), E(X) = \lambda, Var(X) = \lambda$

$X \sim Binomial(n, p) \wedge n$ is large $\wedge p$ is small $\Rightarrow Var(X) \approx np = E(X) \Rightarrow X \sim Poisson(np)$, even if trials are weakly dependent

Poisson X is the number of events that occur in any interval of length t, with the following assumptions:

1. Probability that 1 event occurs in given interval of length h equals $\lambda h + o(h)$, where λ = occurrences per unit time
2. Probability that ≥ 2 events occur in interval of length h equals

$o(h)$, for some $\lim_{h\to 0}\frac{o(h)}{h}=0$

3. For $n, j_1, j_2, \ldots, j_n \in \mathbb{N}$, a set of n nonoverlapping intervals and E_i the event that j_i of the events occur in the i-th interval, $E_1, E_2, \ldots, E_n$ are independent

$X \sim Poisson(\lambda t)$. If $X \sim Poisson(\lambda_1), Y \sim Poisson(\lambda_2) \Rightarrow X+Y \sim Poisson(\lambda_1+\lambda_2)$

4.2 Continuous Random Variables

X is a continuous random variable if $\exists f(x)$ such that

$$\forall x \in \mathbb{R}, f(x) \geq 0 \wedge P(X \in A) = \int_A f(x)dx$$

$f(x)$ is the PDF of X. CDF is defined $\forall a \in \mathbb{R}$,

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

$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{0}^{\infty} P(X>x)dx$ for $X>0$.

For real-valued function $g(x)$, $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$

$Var(X) = E(X-\mu)^2 = E(X^2) - [E(X)]^2$

Uniform X on the segment $[\alpha, \beta]$, PDF is:

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \alpha < x < \beta \\ 0 & \text{otherwise} \end{cases}$$

$X \sim U(\alpha,\beta), E(X) = \frac{\alpha+\beta}{2}, Var(X) = \frac{(\beta-\alpha)^2}{12}$

Note: $(\alpha-\beta)U+\alpha \sim U(\alpha,\beta)$, $F(X) \sim U(0,1)$, $F^{-1}(U)$ is a random variable with cdf $F(x)$

Normal has PDF on two parameters $\mu \in \mathbb{R}, \sigma > 0$:

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

$X \sim N(\mu,\sigma^2), E(X)=\mu, Var(X)=\sigma^2$

$Z \sim N(0,1)$, CDF of Z denoted by $\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}\,dy, \phi(x) = \frac{d}{dx}[\Phi(x)]$

Note: $X \sim N(\mu,\sigma^2) \Rightarrow \frac{X-\mu}{\sigma} \sim N(0,1)$

$Y \sim N(\mu,\sigma^2) \Rightarrow F_Y(a) \sim \Phi(\frac{a-\mu}{\sigma})$

For independent $X_i \sim N(\mu_i, \sigma_i^n)$,

$\sum_{i=1}^{n} X_i \sim N(\sum_{i=1}^{n}, \sum_{i=1}^{n} \sigma_i^2)$

Demoivre-Laplace Limit:

$S_n \sim Bin(n,p) \Rightarrow \frac{S_n - np}{\sqrt{np(1-p)}} \sim Z, np(1-p) \geq 10$

Gamma PDF on parameters $\alpha > 0, \lambda > 0$:

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$\forall x > 1, \Gamma(x) = \int_{0}^{\infty} u^{x-1}e^{-u}\,du = (x-1)\Gamma(x-1)$

$X \sim Gamma(\alpha,\lambda), E(X) = \frac{\alpha}{\lambda}, Var(X) = \frac{\alpha}{\lambda^2}$

For independent random variables:

- $X_i \sim \Gamma(\alpha_i,\lambda) \Rightarrow \sum_{i=1}^{n} X_i \sim \Gamma(\sum_{i=1}^{n}\alpha_i, \lambda)$
- $Z_i \sim N(0,1) \Rightarrow \sum_{i=1}^{n} Z_i^2 \sim \Gamma(\frac{n}{2},\frac{1}{2}) = \chi_n^2$
- $X_i \sim Exp(\lambda) \Rightarrow \sum_{i=1}^{n} X_i \sim \Gamma(n,\lambda)$

Notably, the **Exponential** distribution is the case where $\alpha = 1$

$F(a) = P\{X \leq a\} = 1 - e^{-\lambda a}, a \geq 0$

$X \sim Exp(\lambda), E(X) = \frac{1}{\lambda}, Var(X) = \frac{1}{\lambda^2}$

The uniquely memoryless distribution (Discrete version–Geometric):

$P(X > s+t | X > t) = P(X > s) \forall s,t > 0$

Beta PDF on parameters $a > 0, b > 0$:

$$f(u) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}u^{a-1}(1-u)^{b-1}, u \in [0,1]$$

$X \sim Beta(a,b), E(X) = \frac{a}{a+b}$,

$Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$

$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 x^{a-1}(1-x)^{b-1}$

Note: $Beta(1,1) = U(0,1)$

Cauchy PDF on parameter $\theta \in \mathbb{R}$:

$$f(x) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}, x \in \mathbb{R}$$

$X \sim Cauchy(\theta), \forall x \in \mathbb{N}, E(X^n)$ does not exist.

5 Random Multi-Variables

Joint Distribution Functions

Joint CDF $F(x,y) = P(X \leq x, Y \leq y)$

Marginal CDF $F_X(x) = \lim_{y\to\infty} F(x,y)$

For **discrete** random variables X and Y, $\forall (x,y) \in \mathbb{Z}^2$

Joint PMF $p(x,y) = P(X=x, Y=y)$

Marginal PMF

$$P(X=x) = \sum_{\{y:P(X=x,Y=y)>0\}} P(X=x,Y=y)$$

Conditional Probability of X given that $Y=y, p_Y(y) > 0$

$$p_{X|Y}(x|y) = P(X=x|Y=y) = \frac{p(x,y)}{p_Y(y)}$$

$p_X(x) = \sum_y p_{X|Y}(x|y)p_Y(y)$

$F_{X|Y}(x|y) = P(X \leq x | Y=y) = \sum_{a\leq x} p_{X|Y}(a|y)$

For **jointly continuous** random variables $X, Y, \forall (x,y) \in \mathbb{R}^2$

$\exists f(x,y) > 0, \forall C \subseteq \mathbb{R}^2$

$$P[(X,Y) \in C] = \iint_C f(x,y)\,dx\,dy$$

Joint PDF $f(x,y) = \frac{\partial^2}{\partial x \partial y}F(x,y)$

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

Conditional Probability of X given $Y=y, f_Y(y) > 0$

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

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

$F_{X|Y}(a|y) = P(X \leq a | Y=y) = \int_{-\infty}^{a} f_{X|Y}(x|y)\,dx$

Independence $\iff \forall A_1, A_2, \ldots, A_n \subseteq \mathbb{R}$

$$\begin{aligned} &\iff P(X_1 \in A_1, \ldots, X_n \in A_n) \\ &= \prod_{i=1}^{n} P(X_i \in A_i) \\ &\iff f_{X|Y}(x|y) = f_X(x), \forall x,y \end{aligned}$$

For functions $U = g_1(X,Y), V = g_2(X,Y)$ of jointly continuous X,Y with $f_{XY}(x,y)$, if:

- Transformation is uniquely invertible $x = h_1(u,v), y = h_2(u,v)$
- g_1, g_2 have continuous partial derivatives
- $\forall x,y$ the Jacobian Determinant $J(x,y) \neq 0$

$$J(x,y) = \begin{vmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{vmatrix} = (\frac{\partial g_1}{\partial x})(\frac{\partial g_2}{\partial y}) - (\frac{\partial g_1}{\partial y})(\frac{\partial g_2}{\partial x})$$

$$f_{UV}(u,v) = f_{XY}(x,y)\frac{1}{|J(x,y)|}$$

EXTENDABLE TO >2 VARIABLESAAAAAAAAAAAAA

6 Properties of Expectation

Expectation for Joint Variables

Suppose jointly distributed random variables $X_1, \ldots, X_n$ and $Y = g(X_1, \ldots, X_n)$,

$E(Y) = E[g(X_1, \ldots, X_n)]$

Corollaries:

- Often can decompose X into a sum of n individual events. IMPT! $E(X_1+X_2+\cdots+X_n) = E(X_1)+\cdots+E(X_n)$
- For independent variables and fixed functions g and h, $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$

Co's

Covariance of jointly distributed random variables X and Y is

$$\begin{aligned} Cov(X,Y) =& E[(X-\mu_X)(Y-\mu_Y)] \\ =& E(XY) - E(X)E(Y) \end{aligned}$$

$Cov(X,Y) < 0 \Rightarrow X$ and Y are inversely related X and Y are independent $\Rightarrow Cov(X,Y) = 0$

X and Y are independent $\not\Leftarrow Cov(X,Y) = 0$ *Results*:

- $Cov(X,Y) = Cov(Y,X)$
- $Cov(X,X) = Var(X)$
- $Cov(aX,bY) = abCov(X,Y)$
- $Cov(X,Y+Z) = Cov(X,Y) + Cov(X,Z)$
- $Cov\left(\sum_{i=1}^{n} b_iX_i, \sum_{j=1}^{m} d_jY_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} b_id_jCov(X_i,Y_j)$

Correlation of X and Y is defined when

$\exists Var(X) \neq 0, Var(Y) \neq 0, Cov(X,Y)$

$$\rho = \rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

$-1 \leq \rho \leq 1$, with equality $\iff \exists a,b \in \mathbb{R}, P(Y = a+bX) = 1$

Correlation is the measure of *degree of linearity* between X and Y. Intuitively, the standardised covariance.

$\rho(X,Y) \to \pm 1 \Rightarrow$ high degree of linearity. $\rho(X,Y) = 0 \Rightarrow$ linearity is absent.

For indicator variables I_A, I_B, $Cov(I_A, I_B) = P(B)[P(A|B) - P(A)]$

Conditional Expectation of Y given $X = x$ is

$$\begin{aligned} E(Y|X=x) =& \sum_y p_{Y|X}(y|x) \\ =& \int_y yf_{Y|X}(y|x)\,dy \end{aligned}$$

Alternatively, it is the expectation on a reduced sample space consisting of outcomes for which $Y = y$

Conditional Variance of Y given X is

$$\begin{aligned} Var(Y|X=x) =& E[(Y-E(Y))^2|X=x] \\ =& E(Y^2|X=x) - [E(Y|X=x)]^2 \\ Var(Y|X) =& E[(Y-E(Y))^2|X] \\ =& E(Y^2|X) - [E(Y|X)]^2 \end{aligned}$$

$E(Y|X), Var(Y|X)$ are functions of the random variable X, and hence also random variables.

Expectations!! Variances and Probabilities

- Law of Total Expectation

$$\begin{aligned} E(Y) =& E[E(Y|X)] \\ =& \sum_x E(Y|X=x)p_X(x) \\ =& \sum_x \left[\sum_y yp_{Y|X}(y|x)\right]p_X(x) \\ =& \int_x E(Y|X=x)f_X(x)\,dx \\ =& \int_x \left[\int_y yf_{Y|X}(y|x)\,dy\right]f_X(x)\,dx \end{aligned}$$

- $Var(Y) = Var[E(Y|X)] + E[Var(Y|X)]$
- $P(E) = E[P(E|Y)]$

Moment Generating Functions

MGF of random variable X is $M(t) = E(e^{tX})$, if defined.

$$M(t) = \sum_x e^{tx}p(x) = \int_{-\infty}^{\infty} e^{tx}f(x)\,dx$$

$M^{(r)}$ continuous $\Rightarrow M^{(r)}(0) = \frac{d^r}{dt^r}M(t)\Big|_{t=0} = E(X^r)$

$X \sim Bin(n,p), M(t) = [pe^t + (1-p)]^n$

$X \sim Poisson(n,p), M(t) = e^{\lambda(e^t-1)}$

$X \sim Gamma(\alpha,\lambda), M(t) = \lambda^\alpha(\lambda-t)^{-\alpha}$

$X \sim N(\mu,\sigma^2), M(t) = e^{\mu t+\sigma^2t^2/2}$

$Z = X+Y, M_Z(t) = M_X(t)M_Y(t)$ where mgf's exist

If mgf exists for $a < 0 < b$, it uniquely determines pdf

$Y = a+bX \Rightarrow M_Y(t) = e^{at}M_X(bt)$

$M_{XY}(s,t) = E(e^{sX+tY})$. $M_X(s) = M_{XY}(s,0)$. $M_{XY}(s,t) = M_X(s)M_Y(t)$.

X and Y are independent $\iff M_{XY}(s,t) = M_X(s)M_Y(t)$

7 Limit Theorems

- Markov's Inequality [$E(X)$ exists, and known], $\forall a > 0$ $P(X \geq 0) = 1 \Rightarrow P(X \geq a) \leq \frac{E(X)}{a}$
- Chebyshev's Inequality [Tighter, $\mu, Var(X)$ needed], $\forall k > 0$ $P(|X-\mu| > k) \leq \frac{\sigma^2}{k^2}$
- Chernoff Bounds $P(X \geq a) \leq e^{-ta}M_X(t), \forall t > 0$ $P(X \leq a) \leq e^{-ta}M_X(t), \forall t < 0$
- Jensen's Inequality for convex $f: \forall x, f''(x) \geq 0$ $E(f(X)) \geq f(E(X))$
- For i.i.d random variables $X_1, X_2, \ldots, \forall \epsilon > 0$, $P(|\frac{X_1+\cdots+X_n}{n} - \mu| \geq \epsilon) \to 0$ as $n \to \infty$
- With probability 1, $\frac{X_1+\cdots+X_n}{n} \to \mu$ as $n \to \infty$
- For $Z_1, Z_2, \ldots$ with cdf, mgf $F_{Z_n}(t), M_{Z_n}(t) \wedge \exists Z$ with $F_Z(t), M_Z(t)$, then where $F_Z(t)$ is continuous, $\forall t, M_{Z_n}(t) \to M_Z(t) \Rightarrow F_{Z_n}(t) \to F_Z(t)$

Central Limit Theorem

For i.i.d $X_1, X_2, \ldots$ where $E(X_i) = \mu, Var(X_i) = \sigma^2$, $\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \sim N(0,1)$ as $n \to \infty$