1 Formulas

Formulas

- Permutation:
$$\frac{n!}{(n-r)!}$$
- Combinations:
$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$
- Geometric Series:
$$\sum_{k=0}^{\infty} ar^k = \frac{a}{1-r}$$
- Infinity Series:
$$\sum_{k=0}^{\infty} \frac{z^k}{k!} = e^z$$
- Exponential Result:
$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$$

2 Basic Properties

- PMF (Probability Mass Function) for Discrete RV:
  - $p(k) \geq 0$ for all k
  - $\sum_i p(k_i) = 1$
- PDF (Probability Density Function) for Continuous RV:
  - $f(x) \geq 0$ for all x
  - $\int_{\infty}^{\infty} f(x)\,dx = 1$
- CDF (Cumulative Distribution Function):
  - Non-decreasing function.
  - $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$

Joint Distribution

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

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

Marginal PMF:
$$p_X(x) = \sum_y p(x,y)$$
$$p_Y(y) = \sum_x p(x,y)$$

- $p(x,y) \geq 0$ for all x, y
- $\sum_x \sum_y p(x,y) = 1$

Joint PDF:
$$f(x,y) = P(X=x, Y=y)$$

Marginal PDF:
$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

- $f(x,y) \geq 0$ for all x, y
- $\int_x \int_y p(x,y) \partial y \partial x = 1$

3 Conditional

Conditional Probability for Event

- Definition:
$$P(A|B) = \frac{P(A\cap B)}{P(B)}$$
- Mutiplicative Law:
$$P(A\cap B) = P(B \cap A) = P(A|B)P(B)$$
- Law of Total Probability, Union of all B_i is the Ω
$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

Conditional Probability for Multivariate

- Definition:
$$p(X=x|Y=y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} \qquad f(X=x|Y=y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$
- If X and Y are independent, then their margin PMF/PDF can factor into the product of their Marginal, and canceled by the denominator, thus we got:
$$p(X=x|Y=y) = p(X=x) \qquad f(X=x|Y=y) = f(X=x)$$
- Mutiplication Law:
$$p_{XY}(x,y) = p_{X|Y}(x|y)p_Y(y) \qquad f_{XY}(x,y) = f_{X|Y}(x|y)f_Y(y)$$
- Law of Total Probability:
$$p_X(x) = \sum_y p(x,y) = \sum_y p_{X|Y}(x|y)p_Y(y)$$
$$f_X(x) = \int_{-\infty}^{\infty} f(x,y) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)$$

Bayes's Rule

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

4 Independence

Independence

- RV X and Y are independent iff:
$$\begin{matrix} f(x,y) = f_X(x)f_Y(y) \\ p(x,y) = p_X(x)p_Y(y) \end{matrix} \quad or \quad M_{X,Y}(x,y) = M_X(x)M_Y(y)$$
- Event A and B are independent iff:
$$P(A\cup B) = P(A)P(B)$$

5 Transformation

Normal Transformation (Normalize)

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

Direct Transformation Method

Given $f_X(x)$, find $F_X(x)$, then construct $F_Y(y)$ in terms of $F_X(y)$, next we find the derivative of $F_Y(y)$ to find $f_Y(y)$.

Monotone Transformation Method

If $Y = g(X)$, where g is differentiable and strictly monotonic on some interval I, then the PMF/PDF of Y is given as:
$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|$$

Probability Integral Transformation

If $Z = F_X(X)$ then $Z \sim Uni(0,1)$.

Inverse Integral Transformation

If $X = F^{-1}(U)$ where $U \sim Uni(0,1)$ then X has PDF $F(x)$.

Convolution Method (Sum of two RV)

Given $Z = X+Y$, then we have:
$$p_Z(z) = \sum_x p(x, z-x) \qquad f_Z(z) = \int_{-\infty}^{\infty} f(x, z-x)\,dx$$

If X and Y are independent, we got:
$$p_Z(z) = \sum_x p_X(x)p_Y(z-x) \qquad f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\,dx$$

Bivariate Transformation Method

Suppose X and Y continuous RV and independent, we have two RV, defined as transformation of X and Y, we first use U, V represent X, Y:
$$U = g_1(X,Y) \quad V = g_2(X,Y) \quad \text{and} \quad X = h_1(U,V) \quad Y = h_2(U,V)$$

Then the joint PDF of U and V is given as:
$$f_{U,V}(u,v) = f_{X,Y}(h_1(u,v), h_2(u,v))\,|\det(J(u,v))|$$

Where $J(u,v)$ is the Jacobian Matrix, defined as:
$$J(u,v) = \begin{bmatrix} \frac{\partial h_1}{\partial u} & \frac{\partial h_1}{\partial v} \\ \frac{\partial h_2}{\partial u} & \frac{\partial h_2}{\partial v} \end{bmatrix}$$

And the determinate is calculated as:
$$\frac{\partial h_1}{\partial u} \times \frac{\partial h_2}{\partial v} - \frac{\partial h_1}{\partial v} \times \frac{\partial h_2}{\partial u}$$

6 Expected Value

$$E(X) = \sum_x xp(x) \qquad E(X) = \int_{-\infty}^{\infty} xf(x)\,dx$$

Properties of Expected Value

- Expectation of a constant:
$$E(E(X)) = E(X)$$
- Expectation of Linear Combinations of RV:
$$E\left(a + \sum_{i=1}^{n} b_iX_i\right) = a + \sum_{i=1}^{n} b_iE(X_i)$$

  Especially:
$$E(aX+b) = aE(X)+b$$

  When $a=0$, $b=1$ we got:
$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$
- Notice:
$$E(g(X)) \neq g(E(X))$$
- Expectation of product of RV, If X and Y are independent:
$$E(XY) = E(X)E(Y)$$
- Expected Value of a function of RV, suppose $Y = g(X)$:
$$E(Y) = \sum_x g(x)p(x) \qquad E(Y) = \int_{-\infty}^{\infty} g(x)f(x)\,dx$$

  Especially:
$$E(X^2) = \sum_x x^2p(x) \qquad E(x^2) = \int_{-\infty}^{\infty} x^2f(x)\,dx$$

Conditional Expectation

- Expectation of Y given $X = x$ (fixed), and $h(Y)$ is a function of Y:
$$E(h(Y)|X=x) = \sum_{y_i} h(y_i)p(Y=y_i|X=x)$$
$$E(h(Y)|X=x) = \int_{-\infty}^{\infty} h(y_i)p(Y=y_i|X=x)$$
- Law of total Expectation:
$$E(Y) = E_X(E(Y|X))$$

  The key here is $E(Y|X)$ is a function of X.

7 Variance

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

Standard Deviation

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

Properties of Variance

- Alternative Variance Form:
$$Var(X) = E(X^2) - [E(X)]^2$$
- Variance of sum of RV:
$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$
$$Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$$

  If X and Y are independent, the covariance term is 0, thus:
$$Var(X+Y) = Var(X-Y) = Var(X) + Var(Y)$$

  In General:
$$Var\left(a + \sum_{i=1}^{n} b_iX_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} b_ib_jCov(X_i, X_j)$$

  Especially:
$$Var(aX+b) = a^2Var(X)$$
- If all X_i are mutually independent, then:
$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$
- Variance of product of RV:
$$Var(XY) = E(X^2Y^2) - [E(XY)]^2$$

Law of total Variance

$$Var(Y) = Var(E(Y|X)) + E(Var(Y|X))$$

Covariance

- The Covariance of X and Y is defined as:
$$Cov(X,Y) = E((X-E(X)) * (Y-E(Y)))$$

  Notice that covariance can be positive or negative, contrast to variance which can only take positive value.
- Alternative Covariance Form:
$$Cov(X,Y) = E(XY) - E(X)E(Y)$$
- Property 1:
$$Cov(a+X, Y) = Cov(X,Y)$$
- Property 2:
$$Cov(aX, bY) = abCov(X,Y)$$
- Property 3:
$$Cov(X, Y+Z) = Cov(X,Y) + Cov(X,Z)$$
- Property 4:
$$Cov(aX+bW, cY+dZ) = \begin{matrix} ac*Cov(X,Y) + ad*Cov(X,Z) + \\ bc*Cov(W,Y) + bd*Cov(W,Z) \end{matrix}$$

  In General, if $U = a + \sum_{i=1}^{n} b_iX_i$, $V = c + \sum_{i=1}^{m} d_iX_i$, we have:
$$Cov(U,V) = \sum_{i=1}^{n}\sum_{j=1}^{m} b_id_jCov(X_i, Y_j)$$
- Property 5:
$$Cov(X,X) = Var(X)$$
- Property 6: If X and Y are independent then:
$$Cov(X,Y) = 0$$

  But $Cov(X,Y) = 0$ <u>can't</u> gives us X and Y independent.

Correlation

- The Correlation of X and Y is defined as:
$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \qquad -1 \leq \rho \leq 1$$
- When ρ is close to 1, then X and Y are positively associated.
- When ρ is close to -1, then X and Y are negatively associated.
- When ρ is equals to 0, then X and Y are not associated.

8 Markov and Chebyshev

Markov's Inequality

If X only defined on non negative values, then:
$$P(X \geq t) \leq \frac{E(X)}{t}$$

Chebyshev's Inequality

Let μ and σ^2 be the mean and variance, then for $t > 0$, we set $t = k\sigma$:
$$P(|X-\mu| > t) \leq \frac{\sigma^2}{t^2} \qquad P(|X-\mu| > k\sigma) \leq \frac{1}{k^2}$$

9 Moment Generating Function

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

Properties of Moment Generating Function

- MGF is unique for a distribution, so can prove the distribution that an RV follows.
- MGF can be used to calculate some form of Expectation. That is, the rth moment is:
$$E(X^r) = M^{(r)}(0)$$

  So Variance can also be calculated as the second moment of X subtract the square of the first moment of X, that is:
$$Var(X) = M^{(2)}(0) - [M^{(1)}(0)]^2$$
- rth central moment is defined as:
$$E([X-E(X)])^r$$
- MGF of a transformed function is:
$$M_{aX+b}(t) = e^{bt}M_X(at)$$
- If X and Y are independent RV, then:
$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

10 Gamma Function

$$\Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1}e^{-u}\,du$$

- when α is fraction : $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$
- when α is integer : $\Gamma(\alpha+1) = \alpha!$

a	$1/2$	$3/2$	1	2	3	4
$\Gamma(a)$	$\sqrt{\pi}$	$\sqrt{\pi}/2$	1	1	2	6

11 Law of Large Number (LLN)

Let $X_1, X_2, \ldots$ be independent RV, and $E(X_i) = \mu$, $Var(X_i) = \sigma^2$ (we only require the variance is finite), Let: $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ then for any $\varepsilon > 0$:
$$P\left(\left|\bar{X}_n - \mu\right| \geq \varepsilon\right) \to 0 \qquad \text{as} \quad n \to \infty$$

That is when $n \to \infty$, then the sample mean <u>convergence in probability</u> to the true mean $\bar{X}_n \to \mu$.

12 Monte Carlo Integration

We want to find:
$$I(g) = \int_0^1 g(x)\,dx$$

We first generate n RV $X_1, \ldots X_n$ from $Uni(0,1)$, then we have:
$$\bar{X}_n = \hat{I}(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_i) \qquad \text{as} \quad n \to \infty$$

The key here is that the Uniform distribution on $[0,1]$ has PDF 1, thus the expectation of sample mean is just the function we want.

13 Convergence in Distribution

Let $X_1, \ldots, X_n$ be independent RV. Let X be RV. Then X_n convergence in distribution to X if:
$$\lim_{n\to\infty} F_n(x) = F(x)$$

At all points which F is continuous. We often use MGF to prove convergence in distribution, that is:
$$\lim_{n\to\infty} M_n(t) = M(t) \implies \lim_{n\to\infty} F_n(x) = F(x)$$

For t in an open interval containing zero.

We can use Standard Normal to approximate Poisson, when λ gets large enough, but we first need to Standardiz the Poisson.

14 Central Limit Theorem (CLT)

Let $X_1, X_2, \ldots$ be independent RV, and $E(X_i) = 0$, $Var(X_i) = \sigma^2$ and common CDF/PDF and MGF defined in a neighbourhood of zero.

- If we want Sum of X_i, we define:
$$S_n = \sum_{i=1}^{n} X_i$$

  Then we have:
$$\lim_{n\to\infty} P\left(\frac{S_n}{\sigma\sqrt{n}} \leq x\right) = \Phi(x)$$

  So, we have convergence in distribution of:
$$S_n \to N(0, n\sigma^2)$$
- If we want Average of X_i, we define:
$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  Then we have:
$$\lim_{n\to\infty} P\left(\frac{\bar{X}_n}{\sigma/\sqrt{n}} \leq x\right) = \Phi(x)$$

  So, we have convergence in distribution of:
$$\bar{X}_n \to N(0, \sigma^2/n)$$

If we don't have Expected value 0, we can subtract off the mean and shift the distribution to make it have an expected value of 0.

Normal Approximation

In practise, we can normalize the sum to make it a standard normal:

$$\frac{S - E(S)}{Std(S)} \sim N(0,1)$$

- Binomial: Let $X_1, \ldots, X_n$ be RV that follows Bernoulli distribution with parameter p. So their sum:

  $$S_n = \sum_{i=1}^{n} X_i$$

  follows a Binomial distribution, that is: $S_n \sim Bin(n,p)$, we have that:

  $$Z_n = \frac{S_n - np}{\sqrt{np(1-p)}}$$

  Where $Z_n \sim N(0,1)$.