1. Random Variables

Cumulative Distribution Function (c.d.f.)

$$F_X(x) = P(X \in (-\infty, x]) = P(X \leq x)$$

Probability Mass Function (p.m.f.)

$$f_X(x) = P(X = x)$$

Probability Density Function (p.d.f)

$$f_X(x) = \frac{dF_X(x)}{dx}$$

image.png

- $\int_{-\infty}^{\infty} f_X(x)\ dx = 1$
- $P(a < x < b) = \int_a^b f_X(x)$

Expectation

$$E[X] = \begin{cases} \Sigma_i\ x_i \cdot f_X(x_i) & \text{if discrete} \\ \int_{-\infty}^{\infty} x \cdot f_X(x)\ dx & \text{if continuous} \end{cases}$$

$$E[g(X)] = \begin{cases} \Sigma_i\ g(x_i)_X(x_i) & \text{if discrete} \\ \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\ dx & \text{if continuous} \end{cases}$$

- Comparison: if $P(X \geq a) = 1$, then $E[X] \geq a$
- Linearity: $E[aX + b] = aE[X] + b$
- $E[XY] = E[X] \cdot E[Y]$ if X and Y are independent

Variance

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

- $Var(aX + b) = a^2 \cdot Var(X)$
- $Var(X + Y) = Var(X) + Var(Y)$ if X and Y are independent

2. Discrete Prob Distributions

Bernoulli Distribution

$$X \sim Ber(p) \leftrightarrow P(X = 1) = p \wedge P(X = 0) = 1 - p$$

Interpretation: 1 trial with success or failure

- Mean: $E[X] = p$
- Variance: $Var(X) = p(1 - p)$

Binomial Distribution

$$Y = X_1 + X_2 + ... + X_n$$
$$Y \sim Bin(n, p) \leftrightarrow P(Y = k) = \binom{n}{k}\ p^k(1 - p)^{n-k}$$

Interpretation: number of successes in n trials

- Mean: $E[X] = np$
- Variance: $Var(X) = np(1 - p)$

Poisson Distribution

$$X \sim Pois(\lambda) \leftrightarrow P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

$$e^{\lambda} = \Sigma_{k=0}^{\infty}\frac{\lambda^k}{k!} = 1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{6} + ...$$

Poisson Limit Theorem

Let $Y_n \sim Bin(n, \frac{\lambda}{n})$, $X \sim Pois(\lambda)$. Then $lim_{n\to\infty}P(Y_n = k) = P(X = k)$

- Mean: $E[X] = \lambda$
- Variance: $Var(X) = \lambda$

Discrete Uniform Distribution

$$P(X = x_i) = \frac{1}{n}$$

- Mean: $E[X] = \frac{1}{n}\Sigma_{i=1}^{n}x_i$
- Variance: $Var(X) = E[X^2] - E[X]^2 = \frac{1}{n}\Sigma_{i=1}^{n}x_i^2 - (\frac{1}{n}\Sigma_{i=1}^{n}x_i)^2$

Geometric Distribution

$$X \sim Geom(p) \leftrightarrow P(X = k) = p(1 - p)^{k-1}$$
$$\text{Tail probability: } P(X \geq k) = (1 - p)^{k-1}$$

Interpretation: waiting time for 1st success

- Memoryless: $P(X = i + k \mid X = k) = P(X = i)$
- Mean: $E[X] = \frac{1}{p}$
- Variance: $Var(X) = \frac{1-p}{p^2}$

Negative Binomial Distribution

(don't remember me)

$$X \sim NB(r, p) \leftrightarrow P(X = n) = \binom{n-1}{r-1}\ p^r(1 - p)^{n-r}$$

Sum of r iid Geom(p) random variables

Interpretation: waiting time for r successes

- Mean: $E[X] = \frac{r}{p}$
- Variance: $Var(X) = \frac{r(1-p)}{p^2}$

Hypergeometric Distribution

(don't remember me)

$$X \sim H(n, N, m) \leftrightarrow P(X = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$$

Interpretation: number of successes in choosing n of N, where m of N give success and *(N-m)* of N give failure (without replacement!)

- Mean: $E[X] = n \cdot \frac{m}{N}$
- Variance: $Var(X) = \frac{N-n}{N-1} \cdot n \cdot \frac{m}{N}(1 - \frac{m}{N})$

3. Continuous Prob Distributions

Uniform Distribution

$$X \sim U(a, b) \ \leftrightarrow\ f_X(x) = \frac{1}{b-a}1_{[a,b]}(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in (a, b) \\ 0 & \text{if } x \notin (a, b) \end{cases}$$

- Mean: $E[X] = \frac{a+b}{2}$
- Variance: $Var(X) = \frac{(b-a)^2}{12}$

Exponential Distribution (rmb pdf)

$$X \sim Exp(\lambda) \ \leftrightarrow\ f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

Intepretation: waiting time for one event to happen

- λ is rate at which 'clock' ticks
- Similar to *geometric distribution* (limit as time interval $\to 0$)
- Memoryless: $P(X \geq s + t \mid X \geq t) = P(X \geq s)$
- Mean: $E[X] = \frac{1}{\lambda}$
- Variance: $Var(X) = \frac{1}{\lambda^2}$
- Tail probability: $P(X > t) = e^{-\lambda t}$

Gamma Distribution (no rmb pdf)

$$X \sim \Gamma(\alpha, \lambda) \ \leftrightarrow\ f_X(x) = \begin{cases} \frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Interpretation: waiting time for α events to happen

- $\Gamma(x) = \int_0^{\infty} x^{\alpha-1}e^{-x}\ dx$
- Sum of α independent $Exp(\lambda)$ RVs
- Similar to negative binomial
- Mean: $E[X] = \frac{\alpha}{\lambda}$
- Variance: $Var(X) = \frac{\alpha}{\lambda^2}$

Normal Distribution (rmb pdf!)

$$X \sim N(\mu, \sigma^2) \ \leftrightarrow\ f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
$$Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$$

- Mean: $E[X] = \mu$
- Variance: $Var(X) = \sigma^2$

4. Approximations

Binomial $\to$ Normal

Criteria: $Var(X) = np(1 - p)$ is large enough (say $\geq$10)

Binomial $\to$ Poisson

Criteria: $p \approx 0$ and n is large such that $np \approx \lambda$ and $np(1 - p) \approx \lambda$

Any iid → Normal

Central Limit Theorem: with large n, sum of iids gives normal distribution

Continuity Correction (to Normal)

Let X be discrete, let Y be normal approximation.

$P(X = k) \approx P(k - \frac{1}{2} < Y < k + \frac{1}{2})$

(If you're taking say $2X$, you need double the continuity correction!)

5. Function of Random Variables

Question: What is the distribution of $Y = g(X)$ given X and g?

Procedure: (Idea is to find $F_y(y)$ and then $f_Y(y)$, which describes the distribution)

- $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq h(y)) = \dots$ (integral of $f_X(x)$ over some interval)
- $f_Y(y) = \frac{d}{dy} F_Y(y) = \dots$

(See more examples!)

6. Joint Probability: Marginal and Conditional

Joint p.m.f

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

Joint p.d.f

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

Marginal p.m.f

Sum joint p.m.f over all possible values of y

$$f_X(x) = P(X = x) = \Sigma_y f_{X,Y}(x,y)$$

Marginal p.d.f

Integrate joint p.d.f over all possible values of y

$$f_X(x) = \int f_{X,Y}(x,y)\ dy$$

Conditional p.m.f and p.d.f

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

Independence of RVs

X and Y are independent $\leftrightarrow f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$

Expectation of function of RVs

Discrete

$E[g(X,Y)] = \Sigma_(x,y) g(x,y) \cdot f_{X,Y}(x,y)$

Continuous

$E[g(X,Y)] = \int \int g(x,y) \cdot f_{X,Y}(x,y)\ dx\ dy$

7. Covariance

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

To find $Cov(X,Y)$:

- Find $E[X]$, $E[Y]$, $E[XY]$
- Use $Cov(X,Y) = E[XY] - E[X] \cdot E[Y]$

Properties of covariance

- $Cov(X,Y) = E[XY]$ if $E[X] = 0$ or $E[Y] = 0$
- $Cov(X,Y) = 0$ if X and Y are independent
- $Cov(aX + \alpha, bY + \beta) = ab \cdot Cov(X,Y)$
- $Cov(X,X) = Var(X)$

Sum of Variance of RVs

$$Var(X_1 + ... + X_n) = \Sigma_{i=1}^n Var(X_i) + 2 \cdot \Sigma_{1 \leq i < j \leq n} Cov(X_i, X_j)$$

Correlation Coefficient

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

Properties of correlation coefficient

- $\rho(X,Y) = Cov(\tilde{X}, \tilde{Y})$ where $\tilde{X} = \frac{X - E[X]}{\sigma_X}$
- $\rho(aX + b, cY + d) = \rho(X,Y)$
- $-1 \leq \rho(X,Y) \leq 1$
- $\rho(X,Y) = 1$ if $Y = aX$ for some $a > 0$
- $\rho(X,Y) = -1$ if $Y = aX$ for some $a < 0$

8. Sum of Independent RVs

What's the distribution of $X + Y$?

$f_{X+Y}(z) = (f_x * f_Y)(z) = \int f_X(z - y) \cdot f_Y(y)\ dy = \int f_X(x) \cdot f_Y(z - y)\ dx$

Exponential

$X, Y \sim Exp(\lambda) \rightarrow f_{X+Y}(z) = \lambda^2 z e^{-\lambda z}$ $(z > 0)$

Gamma

$X \sim \Gamma(\alpha, \lambda), Y \sim \Gamma(\beta, \lambda) \rightarrow X + Y \sim \Gamma(\alpha + \beta, \lambda)$

Poisson

$X \sim Pois(\lambda_1), Y \sim Pois(\lambda_2) \rightarrow X + Y \sim Pois(\lambda_1 + \lambda_2)$

9. Multi-Dimensional Change of Variables

Let $\overrightarrow{X} = (X_1, X_2)$ with p.d.f $f_{\overrightarrow{X}}(x_1, x_2)$

Let $\overrightarrow{g} = (g_1, g_2)$ with inverse $\overrightarrow{h} = (h_1, h_2)$

Let $\overrightarrow{Y} = (g_1(\overrightarrow{X}), g_2(\overrightarrow{X}))$

$$f_{\overrightarrow{Y}}(\overrightarrow{y}) = \frac{f_{\overrightarrow{X}}(\overrightarrow{x})}{|J_{\overrightarrow{g}}(\overrightarrow{x})|} = f_{\overrightarrow{X}}(\overrightarrow{x}) |J_{\overrightarrow{h}}(\overrightarrow{y})|$$

$$J_{\overrightarrow{g}}(\overrightarrow{x}) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1}(\overrightarrow{x}) & \frac{\partial g_1}{\partial x_2}(\overrightarrow{x}) \\ \frac{\partial g_2}{\partial x_1}(\overrightarrow{x}) & \frac{\partial g_2}{\partial x_2}(\overrightarrow{x}) \end{vmatrix}$$

Steps for multi-dimensional change of variables

1. Transformation $\overrightarrow{g}$
2. Range of $\overrightarrow{g}$ (and so $\overrightarrow{Y}$)
3. Inverse $\overrightarrow{h}$
4. Jacobian ($J_{\overrightarrow{g}}(\overrightarrow{x})$ or $J_{\overrightarrow{g}}(\overrightarrow{x})$, usually latter)

⇒ Profit! You can use formula to find $f_{\overrightarrow{Y}}(\overrightarrow{y})$.

(We usually substitute $f_{\overrightarrow{X}}(\overrightarrow{x})$ for $f_{\overrightarrow{X}}(\overrightarrow{h}(\overrightarrow{y}))$.)

10. Multivariate Standard Normals

Let $\overrightarrow{X} = (X_1, \dots, X_n)$ be multivariate normal with mean $\overrightarrow{\mu}$ and covariance matrix Σ.

$$f_{\overrightarrow{X}}(\overrightarrow{x}) = \frac{1}{\sqrt{(2\pi)^n det(\Sigma)}} e^{-\frac{1}{2}(\overrightarrow{x} - \overrightarrow{\mu})^T \Sigma^{-1} (\overrightarrow{x} - \overrightarrow{\mu})}$$

Multivariate normal fully determined by $\overrightarrow{\mu}$ and Σ.

- Σ is symmetric and positive-definite
- $f(\overrightarrow{x}) = Ce^{-Q(\overrightarrow{x})}$ where $Q(\overrightarrow{x})$ is some quadratic polynomial

Properties of multivariate normal

- If $\Sigma_{ii} = \sigma_i^2$ and $\Sigma_{ij} = 0$ for $i \neq j$, then $X_1 \dots X_n$ are independent with $X_i \sim N(\mu_i, \sigma_i^2)$
- Affine transformations of i.i.d standard normal → multivariate normal
- Affine transformations of multivariate normal → multivariate normal
- All multivariate normals can be expressed as affine transformations of i.i.d standard normals
- Marginal distributions are also multivariate normals, but of a lower dimension

(More details: see the book.)

11. Bivariate Standard Normals

Helpful things (2x2)

- $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \rightarrow A^{-1} = \frac{1}{detA} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$
- $detA = ad - bc$

(See more examples!)

12. Expectation, Reloaded

$$E[g(\overrightarrow{X})] = \begin{cases} \Sigma_{\overrightarrow{x}} g(\overrightarrow{x}) f(\overrightarrow{x}) & \text{if discrete} \\ \int g(\overrightarrow{x}) f(\overrightarrow{x}) & \text{if continuous} \end{cases}$$

Conditional Expectation

$$E[g(X)|Y = y] = \begin{cases} \Sigma_x g(x) \cdot f_{X|Y}(x|y) & \text{if discrete} \\ \int g(x) \cdot f_{X|Y}(x|y) & \text{if continuous} \end{cases}$$

Properties of conditional expectations

- $E[g(X)|A] = \frac{E[g(X)1_A]}{P(A)}$
- $E[g(X)] = \Sigma E[g(X)|A_i] \cdot P(A_i)$ for some partition $A_i \dots A_n$
- $E[g(X,Y)] = E[E[g(X,Y)|Y]]$ and $P(A) = E[P(A|Y)]$
- $E[g(X)|Y] = E[g(X)]$ if X and Y are independent
- $E[g(X) \cdot h(Y)|X] = g(X) \cdot E[h(Y)|X]$

Conditional Variance

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

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

13. Moment Generating Functions

$$M_X(t) = E[e^{tX}]$$

M_x is well-defined if $M_X(t) < \infty$ $\forall t \in (-\eta, \eta)$ for some $\eta > 0$

$$E[X^k] = M^{(k)}(0) = [\frac{d^k M(t)}{d^k t}]_{t=0}$$

- $E[X] = M'(0)$
- $E[X^2] = M''(0)$

Properties

- $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$ (if X and Y are independent)
- X and Y have the same distribution if $M_X(t) = M_Y(t)$

Common MGFs

Bernoulli	$1 - p + pe^t$
Binomial	$(1 - p + pe^t)^n$
Exponential	$\frac{\lambda}{\lambda - t}$
Poisson	$e^{-\lambda(1-e^t)}$