Probability Reference Sheet

By Danny Liu

1 Probability Basics

Let E be an event and S the sample space.
$\mathbb{P}(E) \geq 0$, $\mathbb{P}(S) = 1$, $\mathbb{P}(E) = 1 - \mathbb{P}(\overline{E})$
If $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$
$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$
$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
For independent events, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ and $\mathbb{P}(A|B) = \mathbb{P}(A)$
For mutually exclusive events, $\mathbb{P}(A \cap B) = 0$ and $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

2 Bayes' Theorem

If a sample space S can be partitioned into disjoint events $S = A_1 \cup A_2 \cup ... \cup A_n$ and $B \subseteq S$, then
$\mathbb{P}(B) = \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \mathbb{P}(B|A_2)\mathbb{P}(A_2) + ... + \mathbb{P}(B|A_n)\mathbb{P}(A_n)$
$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\mathbb{P}(B|A_1)\mathbb{P}(A_1)+...+\mathbb{P}(B|A_n)\mathbb{P}(A_n)}$

3 Expected Value and Variance

$\mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x)dx$
$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$
$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$

4 Covariance and Correlation

$\mathrm{Var}(\sum_i X_i) = \sum_i \mathrm{Var}(X_i) + 2\sum\limits_{i<j} \mathrm{Cov}(X_i, X_j)$
$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = E[XY] - E[X]E[Y]$
$\mathrm{Cov}(a(X_1 + X_2), Y) = a\mathrm{Cov}(X_1 + X_2, Y) = a\mathrm{Cov}(X_1, Y) + a\mathrm{Cov}(X_2, Y)$
If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$
Define correlation of X and Y as $\rho(X, Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{\langle X,Y\rangle}{\|X\|\|Y\|} = \cos\theta$
$\rho(X, Y) = \pm 1 \iff Y = aX + b$ for some $a, b \in \mathbb{F}$
Covariance is an inner product over the quotient space of r.vs with finite second moments.
Standard deviation is the norm on this quotient space.

5 Random Variables

A random variable is a function $X : S \to \mathbb{R}$ where S is the sample space.
For discrete X, Y, the joint pmf is $p_{X,Y}(x, y) = \mathbb{P}[X = x \cap Y = y]$, where $\sum_x \sum_y p_{X,Y}(x, y) = 1$
For continuous X, Y, the joint pdf is $\int \int_{(x,y) \in R} f(x, y)dxdy = \mathbb{P}[(x, y) \in R]$
Note that although $\int_a^b f(x)dx = \mathbb{P}[a \leq X \leq b]$ for continuous X, $\mathbb{P}[X = a] = 0$
If $F(a) = \mathbb{P}[X \leq a] = \int_{-\infty}^a f(x)dx$ is the cdf, then $f(x) = \frac{d}{dx}F(x)$
Continuous (discrete) r.vs are independent if and only if $f_{X,Y}(x, y) = f_X(x)f_Y(y)$
The marginal density of X is $f_X(x) = \int_y f(x, y)dy$

5.1 Conditional Expectation

The conditional distribution of X given Y is $p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{\mathbb{P}[X=x \cap Y=y]}{\mathbb{P}[Y=y]}$
The conditional expectation of X given Y is $\mathbb{E}[X|Y = y] = \sum_x x p_{X|Y}(x|y)$
$\mathbb{E}[\mathbb{E}[X|Y]] = \sum_y \mathbb{E}[X|Y = y]p_Y(y) = \mathbb{E}[X]$

5.2 Functions of Random Variables

Suppose X is a random variable and $Y = g(X)$. If g is increasing and differentiable, then
$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \cdot \frac{d}{dy}g^{-1}(y) & \text{if } g(a) \leq y \leq g(b) \\ 0 & \text{otherwise} \end{cases}$

6 Inequalities and Limit Theorems

6.1 Central Limit Theorem

Let $X_1, X_2, ..., X_n$ be n iid r.vs with mean μ and variance σ^2. Then for n sufficiently large,

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

where $\bar{X} = \frac{X_1+X_2+...+X_n}{n}$ has mean μ and variance $\frac{\sigma^2}{n}$. In practice, this works well for $n \geq 30$.

6.2 Markov's Inequality

Let X be a r.v that takes on only non-negative values. Then for any $a > 0$, $\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$

6.3 Chebyshev's Inequality

Let X be a r.v with mean μ and variance σ^2. Then for any $k > 0$, $\mathbb{P}[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}$

6.4 Law of Large Numbers

Let $X_1, X_2, ..., X_n$ be iid r.vs with mean μ and finite variance. Then $\mathbb{P}\left[\lim_{n\to\infty} \frac{X_1+X_2+...+X_n}{n} = \mu\right] = 1$

7 Moment Generating Functions

$M_X(t) = \mathbb{E}[e^{tX}] = \int_x e^{tx} f(x)dx$, where $\mathbb{E}[X], \mathbb{E}[X^2], \mathbb{E}[X^3], ...$ are the moments of X
Consider Taylor expansion of $e^{tX} = 1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + ...$
If two r.vs have the same mgf for all $t \in \mathbb{R}$, then they have the same distribution for all $t \in \mathbb{R}$.

7.1 Properties of MGF

1. $M_X(t) = \sum_{k=0}^{\infty} \mathbb{E}[x^k] \frac{t^k}{k!}$
2. $\mathbb{E}[X^k] = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0}$ for any integer $k > 0$
3. If X, Y are r.vs and $Y = aX + b$, then $M_Y(t) = e^{bt} M_X(at)$
4. If $X_1, X_2, ..., X_n$ are independent r.vs, then $M_{X_1+X_2+...+X_n}(t) = M_{X_1}(t) M_{X_2}(t) ... M_{X_n}(t)$

8 Common Distributions

8.1 Binomial Distribution

$X \sim Bin(n, p)$ keeps tracks of the number of n Bernoulli trials that ends up being a success.
$p_X(i) = p^i(1-p)^{n-i}\binom{n}{i}$, $M_X(t) = (pe^t + 1 - p)^n$, $\mathbb{E}[X] = np$, $\mathrm{Var}(X) = np(1-p)$

8.2 Poisson Distribution

$X \sim Poi(\lambda)$ approximates binomial well when n is large, p is small and $\lambda = np$ is moderate.
$p_X(i) = e^{-\lambda}\frac{\lambda^i}{i!}$, $M_X(t) = e^{\lambda(e^t-1)}$, $\mathbb{E}[X] = \lambda$ $\mathrm{Var}(X) = \lambda$

8.3 Geometric Distribution

$X \sim Geo(p)$ counts the number of Bernoulli trials needed to get one success.
$p_X(i) = (1-p)^{i-1}p$, $M_X(t) = \frac{pe^t}{1-(1-p)e^t}$, $\mathbb{E}[X] = \frac{1}{p}$, $\mathrm{Var}(X) = \frac{1-p}{p^2}$

8.4 Uniform Distribution

$X \sim U(a, b)$, $\mathbb{E}[X] = \frac{1}{2}(a+b)$, $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$
$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases}$, $M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$

8.5 Exponential Distribution

$X \sim Exp(\lambda)$, $\mathbb{E}[X] = \frac{1}{\lambda}$, $\mathrm{Var}(X) = \frac{1}{\lambda^2}$
$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$, $M_X(t) = \frac{\lambda}{\lambda - t}$
$\mathbb{P}(X > s + t | X > s) = \mathbb{P}(X > t)$ as exponential variables are memoryless and vice-versa.

8.6 Normal Distribution

$X \sim \mathcal{N}(\mu, \sigma^2)$, $\mathbb{E}[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$, $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$, $M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$
If $Z = \frac{X-\mu}{\sigma}$, then $Z \sim \mathcal{N}(0,1)$ and we define $\phi(z) = \mathbb{P}(Z \le z)$, $\phi(z) = 1 - \phi(-z)$
$\mathbb{P}[a \le X \le b] = \mathbb{P}[X \le b] - \mathbb{P}[X \le a] = \phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})$

9 Miscellaneous

The joint $f_{X,Y}(x,y) = g(x^2 + y^2)$ is radial if and only if the marginals are Gaussian.

9.1 Fundamental Theorem of Calculus

Let f be a continuous real-valued function on $[a,b]$ and F be the antiderivative of f.
Then $\int_a^b f(t)dt = F(b) - F(a)$ and $\frac{d}{dx} \int_a^x f(t)dt = f(x)$

9.2 Integrals

$\int udv = uv - \int vdu$, $\int_0^\infty f(x)dx = \lim\limits_{b\to\infty} \int_0^b f(x)dx$
$\int_s^\infty \lambda e^{-\lambda x} dx = e^{-\lambda s}$, $\int xe^{-\lambda x} dx = \frac{1}{\lambda^2}(-\lambda e^{-\lambda x} x - e^{-\lambda x}) + C$
$\int \frac{1}{1+x^2} dx = \arctan(x) + C$, $\int \frac{x^2}{1+x^2} dx = \int 1 - \frac{1}{1+x^2} dx$, $\int \frac{x^3}{1+x^2} dx = \int x - \frac{x}{1+x^2} dx$