A Concise Probability and Statistics Review

Ease of use has been favoured over formality and rigour. In particular, some results or definitions make reference to concepts defined later. This is deliberate and has been done in order to give the material a more compact organisation. In some cases general results are accompanied by more specific results, again for ease of reference.

Giulio Agostini
giulio.agostini@gmail.com

Probability

Notation

$AB \stackrel{\text{def}}{=} A \cap B \qquad \sim A \stackrel{\text{def}}{=} \Omega \setminus A$

Properties of Probability

$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[AB]$

If two events are mutually exclusive (i.e. incompatible, $\mathbb{P}[AB] = 0$)
$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B]$

Conditional probability

$\mathbb{P}[A|B] \stackrel{\text{def}}{=} \frac{\mathbb{P}[AB]}{\mathbb{P}[B]}$

Properties of Conditional probability

$\mathbb{P}[AB] = \mathbb{P}[A|B]\,\mathbb{P}[B]$

If two events are independent ($\mathbb{P}[A|B] = \mathbb{P}[A]$ and $\mathbb{P}[B|A] = \mathbb{P}[B]$)
$\mathbb{P}[AB] = \mathbb{P}[A]\,\mathbb{P}[B]$

$\mathbb{P}[A] = \mathbb{P}[A|B]\,\mathbb{P}[B] + \mathbb{P}[A| \sim B]\,\mathbb{P}[\sim B]$
More generally, if B_j are mutually exclusive
$\mathbb{P}[A] = \sum_{j=1}^{n} \mathbb{P}[A|B_j]\,\mathbb{P}[B_j]$ (Law of alternatives)
Even more generally (making this applicable to continuous distributions as well)
$\mathbb{P}[A] = \mathbb{E}[\mathbb{P}[A|X = x]]$, where X is a random variable.

$\mathbb{P}[B|A] = \frac{\mathbb{P}[A|B]\mathbb{P}[B]}{\mathbb{P}[A]}$
$\mathbb{P}[B|A] = \frac{\mathbb{P}[A|B]\mathbb{P}[B]}{\mathbb{P}[A|B]\mathbb{P}[B]+\mathbb{P}[A|\sim B]\mathbb{P}[\sim B]}$
More generally, if B_j are mutually exclusive
$\mathbb{P}[B|A] = \frac{\mathbb{P}[A|B]\mathbb{P}[B]}{\sum_{j=1}^{n} \mathbb{P}[A|B_j]\mathbb{P}[B_j]}$ (Bayes)

Random variables, expectation, mean, variance, covariance

Independence

X and Y are independent $\stackrel{\text{def}}{\Leftrightarrow} f_{X,Y}(x, y) = f_X(x) f_Y(y)$.

Uncorrelatedness

X and Y are uncorrelated $\overset{\text{def}}{\Leftrightarrow}$ $\mathrm{Cov}\left[X, Y\right] = 0$.

If X and Y are independent, they are uncorrelated.
The converse doesn't hold in general, i.e. $\mathrm{Cov}\left[X, Y\right] = 0$ doesn't necessarily imply independence.
However, if X and Y are normally distributed, $\mathrm{Cov}\left[X, Y\right] = 0$ does imply independence.
(((Does this hold for other distributions?)))

Expectation

$\mathbb{E}\left[X\right] \overset{\text{def}}{=} \sum_j x_j f_X(x_j)$ $\qquad$ $\mathbb{E}\left[X\right] \overset{\text{def}}{=} \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x$

More generally
$\mathbb{E}\left[g(X)\right] \overset{\text{def}}{=} \sum_j g(x_j) f_X(x_j)$ $\qquad$ $\mathbb{E}\left[g(X)\right] \overset{\text{def}}{=} \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x$

Properties of Expectation

$\mathbb{E}\left[\sum_{i=1}^n a_i X_i\right] = \sum_{i=1}^n a_i \mathbb{E}\left[X_i\right]$
$\mathbb{E}\left[XY\right] = \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right] + \mathrm{Cov}\left[X, Y\right]$
$\mathbb{E}\left[X^2\right] = \mu_X^2 + \mathrm{Var}\left[X\right]$
$\mathbb{E}\left[XY\right] \leq \sqrt{\mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]}$ (Cauchy-Schwartz)
If $g(\cdot)$ is convex, then $\mathbb{E}\left[g(X)\right] \geq g(\mathbb{E}\left[X\right])$ (Jensen)

If X and Y are independent
$\mathbb{E}\left[XY\right] = \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right]$
More generally
$\mathbb{E}\left[g_1(X) g_2(Y)\right] = \mathbb{E}\left[g_1(X)\right]\mathbb{E}\left[g_2(Y)\right]$

Mean

$\mu_X \overset{\text{def}}{=} \mathbb{E}\left[X\right]$

Variance

$\mathrm{Var}\left[X\right] \overset{\text{def}}{=} \mathbb{E}\left[(X - \mu_X)^2\right] = \mathbb{E}\left[X^2\right] - \mu_X^2$

Standard deviation

$\sigma_X \overset{\text{def}}{=} \sqrt{\mathrm{Var}\left[X\right]}$

Properties of Variance

$\mathrm{Var}\left[a\right] = 0$
$\mathrm{Var}\left[X + a\right] = \mathrm{Var}\left[X\right]$
$\mathrm{Var}\left[aX\right] = a^2\mathrm{Var}\left[X\right]$
$\mathrm{Var}\left[X + Y\right] = \mathrm{Var}\left[X\right] + \mathrm{Var}\left[Y\right] + 2\mathrm{Cov}\left[X, Y\right]$
$\mathrm{Var}\left[X - Y\right] = \mathrm{Var}\left[X\right] + \mathrm{Var}\left[Y\right] - 2\mathrm{Cov}\left[X, Y\right]$
$\mathrm{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathrm{Var}\left[X_i\right] + 2\sum_i \sum_{j<i} \mathrm{Cov}\left[X_i, X_j\right]$
$\mathrm{Var}\left[\sum_{i=1}^n a_i X_i\right] = \sum_{i=1}^n a_i^2 \mathrm{Var}\left[X_i\right] + 2\sum_i \sum_{j<i} a_i a_j \mathrm{Cov}\left[X_i, X_j\right]$

$$\begin{aligned}\mathrm{Var}[XY] = {} & \mu_Y^2 \mathrm{Var}[X] + \mu_X^2 \mathrm{Var}[Y] + 2\mu_X\mu_Y \mathrm{Cov}[X,Y] + \\ & -(\mathrm{Cov}[X,Y])^2 + \mathbb{E}\left[(X-\mu_X)^2(Y-\mu_Y)^2\right] + \\ & +2\mu_Y \mathbb{E}\left[(X-\mu_X)^2(Y-\mu_Y)\right] + 2\mu_X \mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)^2\right]\end{aligned}$$

$\mathrm{Var}\left[X^2\right] = 4\mu_X^2 \mathrm{Var}[X] - (\mathrm{Var}[X])^2 + \mathbb{E}\left[(X-\mu_X)^4\right] + 4\mu_X \mathbb{E}\left[(X-\mu_X)^3\right]$

If X_j are uncorrelated (or independent, therefore uncorrelated)
$\mathrm{Var}\left[\sum_{i=1}^n a_i X_i\right] = \sum_{i=1}^n a_i^2 \mathrm{Var}[X_i]$

If X and Y are independent
$\mathrm{Var}[XY] = \mu_Y^2 \mathrm{Var}[X] + \mu_X^2 \mathrm{Var}[Y] + \mathrm{Var}[X]\,\mathrm{Var}[Y]$

Covariance

$\mathrm{Cov}[X,Y] \overset{\text{def}}{=} \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] = \mathbb{E}[XY] - \mu_X\mu_Y$

Properties of Covariance

$\mathrm{Cov}[X,X] = \mathrm{Var}[X]$
$\mathrm{Cov}[X,Y] = \mathrm{Cov}[Y,X]$
$\mathrm{Cov}[X,a] = 0$
$\mathrm{Cov}[aX,bY] = ab\mathrm{Cov}[X,Y]$
$\mathrm{Cov}[X_1+X_2,Y] = \mathrm{Cov}[X_1,Y] + \mathrm{Cov}[X_2,Y]$
$\mathrm{Cov}\left[\sum_{i=1}^n a_i X_i, \sum_{i=1}^m b_i Y_i\right] = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathrm{Cov}[X_i,Y_j]$

Correlation

$\rho_{X,Y} = \frac{\mathrm{Cov}[X,Y]}{\sigma_X\sigma_Y}$

Properties of Correlation

$-1 \le \rho_{X,Y} \le 1$

Conditional density function

$f_{Y|X}(y|x) \overset{\text{def}}{=} \frac{f_{X,Y}(x,y)}{f_X(x)}$
That is to say, a normalised version of $f_{X,Y}(x,y)$.

Conditional expectation

$\mathbb{E}[g(X,Y)|X=x] \overset{\text{def}}{=} \sum_j g(x,y_j) f_{Y|X}(y_j|x)$ $\qquad$ $\mathbb{E}[g(X,Y)|X=x] \overset{\text{def}}{=} \int_{-\infty}^{\infty} g(x,y) f_{Y|X}(y|x)\,\mathrm{d}y$

Properties of Conditional expectation

$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X=x]]$
Note that $\mathbb{E}[Y|X=x] = h(x)$. More generally
$\mathbb{E}[g(Y)] = \mathbb{E}[\mathbb{E}[g(Y)|X=x]]$

Estimators

Sample mean

$\bar{X} \overset{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n X_i$

Properties of Sample mean

$\mathbb{E}\left[\bar{X}\right] = \mu_X \qquad \mathrm{Var}\left[\bar{X}\right] = \frac{\mathrm{Var}[X]}{n}$

If X_i are IID with $\mu = \mathbb{E}\left[X_i\right] < \infty$
$\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}|X_i - \mu| > \epsilon\right] = 0 \qquad \forall\epsilon > 0$ (Weak Law of Large Numbers)

If X_i are IID with $\mu = \mathbb{E}\left[X_i\right] < \infty$ and $\sigma^2 = \mathbb{E}\left[(X_i - \mu)^2\right] < \infty$
$\mathbb{P}\left[\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu\right] = 1 \qquad \forall\epsilon > 0$ (Strong Law of Large Numbers)

If X_i are IID and we define $Z_n \overset{\text{def}}{=} \frac{\bar{X}_n - \mathbb{E}\left[\bar{X}_n\right]}{\sigma_{\bar{X}_n}} = \frac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}}$
$\lim_{n\to\infty} F_{Z_n}(z) = \Phi(z)$ (Central Limit Theorem)

Sample variance

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Properties of Sample variance

$\mathbb{E}\left[S^2\right] = \sigma_X^2$

Distribution-independent results

Distribution of minimum and maximum

$Y = \max(X_1, \ldots, X_n) \Rightarrow F_Y(y) = \prod_{i=1}^{n} F_{X_i}(y)$
$Y = \min(X_1, \ldots, X_n) \Rightarrow F_Y(y) = 1 - \prod_{i=1}^{n}(1 - F_{X_i}(y))$

Distribution of sum and difference

$Z = X + Y \Rightarrow f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\,\mathrm{d}x = \int_{-\infty}^{\infty} f_{X,Y}(z - y, y)\,\mathrm{d}y$
$Z = X - Y \Rightarrow f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, x - z)\,\mathrm{d}x = \int_{-\infty}^{\infty} f_{X,Y}(z + y, y)\,\mathrm{d}y$

If X and Y are independent and $Z = X + Y$
$f_Z(z) = \int_{-\infty}^{\infty} f_Y(z - x) f_X(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} f_X(z - y) f_Y(y)\,\mathrm{d}y$
(convolution of $f_X(\cdot)$ and $f_Y(\cdot)$)

Discrete distributions

Properties of Uniform distribution

$\mu_X = (N + 1)/2 \qquad \mathrm{Var}\left[X\right] = (N^2 - 1)/12$

Bernoulli distribution

$f_X(0; p) \overset{\text{def}}{=} 1 - p \qquad f_X(1; p) \overset{\text{def}}{=} p$

Properties of Bernoulli distrbution

The Bernoulli distribution models an experiment with two possible outcomes: failure (0) or success (1).
$\mu_X = p \qquad \mathrm{Var}\left[X\right] = p(1 - p)$

If $X_i \sim \text{Bernoulli}(p)$ and they are independent
$\sum_i X_i \sim \text{Binomial}(p)$

Binomial distribution

$f_X(x; n, p) \overset{\text{def}}{=} \binom{n}{x} p^x (1-p)^{n-x}$

Properties of Binomial distrbution

The Binomial distribution models the number of successes from n independent Bernoulli experiments.
$\mu_X = np \qquad \mathrm{Var}[X] = np(1-p)$

Geometric distribution

$f_X(x; p) \overset{\text{def}}{=} p(1-p)^x$

Properties of Geometric distrbution

The Geometric distribution models the number of failed independent Bernoulli experiments before the first success.
$\mu_X = (1-p)/p \qquad \mathrm{Var}[X] = np(1-p)/p^2$

Negative Binomial distribution

$f_X(x; r, p) \overset{\text{def}}{=} \binom{r+x-1}{x} p^r (1-p)^x$

Properties of Negative Binomial distrbution

The Negative Binomial distribution models the number of failed independent Bernoulli experiments before the r-th success.
Note that for $r = 1$ this is a Geometric distribution.
$\mu_X = np \qquad \mathrm{Var}[X] = np(1-p)$

Poisson distribution

$f_X(x; \lambda) \overset{\text{def}}{=} \frac{e^{-\lambda}\lambda^x}{x!}$

Properties of Poisson distrbution

If $\lambda = \nu t$, the Poisson distribution models the number of events occurring in $(0, t)$ for a Poisson process having average rate ν.
$\mu_X = \lambda \qquad \mathrm{Var}[X] = \lambda$

If $X_i \sim \mathrm{Poisson}(\lambda_i)$ and they are independent
$\sum_i X_i \sim \mathrm{Poisson}(\sum_i \lambda_i)$

Continuous distributions

Properties of Uniform distribution

$\mu_X = (a+b)/2 \qquad \mathrm{Var}[X] = (b-a)^2/12$

Normal distribution

$f_X(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Properties of Normal distribution

$\mu_X = \mu \qquad \mathrm{Var}\left[X\right] = \sigma^2$

If $X \sim N(\mu_X, \sigma_X^2)$
$aX \sim N(a\mu_X, a^2\sigma_X^2)$

If $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$ and they are independent
$X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$
$X - Y \sim N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2)$

More generally, if $X_i \sim N(\mu_i, \sigma_i^2)$ and they are independent
$\sum_i X_i \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$

Even more generally,
$\sum_i a_i X_i \sim N(\sum_{i=1}^n a_i\mu_i, \sum_{i=1}^n a_i^2\sigma_i^2)$

If we drop the independence hypothesis, we still have
$\sum_i a_i X_i \sim N(\cdot, \cdot)$

If $X \sim N(0, 1)$
$\mathbb{E}\left[X^2\right] = 1$
$\mathrm{Var}\left[X^2\right] = 2$
$X^2 \sim \chi_1^2$

More generally, if $X_i \sim N(0, 1)$ are n independent variables and $Y = \sum_{i=1}^n X_i^2$
$\mathbb{E}\left[Y\right] = n$
$\mathrm{Var}\left[Y\right] = 2n$
$Y \sim \chi_n^2$

If $X \sim N(\mu, \sigma^2)$
$\mathbb{E}\left[X^2\right] = \mu^2 + \sigma^2$
$\mathrm{Var}\left[X^2\right] = 4\mu^2\sigma^2 + 3(\sigma^2)^2$

(((Generalise this to X^n (cf. Hunt, Kennedy))))

If $X \sim N(\mu, \sigma^2)$
$\mathbb{E}\left[\mathrm{e}^{aX}\right] = \mathrm{e}^{a\mu + \frac{1}{2}a^2\sigma^2}$
$F_{\mathrm{e}^{aX}}(x) = \mathrm{e}^{a\mu + \frac{1}{2}a^2\sigma^2}\Phi(x; \mu + a\sigma^2, \sigma^2)$

Lognormal distribution

$f_X(x; \mu, \sigma^2) = \frac{1}{x\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \qquad \forall x \geq 0$

Properties of Lognormal distribution

The Lognormal distribution models variables whose natural logarithm is normally distributed with parameters μ and σ^2.
$\mu_X = \mathrm{e}^{\mu + \frac{1}{2}\sigma^2} \qquad \mathrm{Var}\left[X\right] = \mathrm{e}^{2\mu + 2\sigma^2} - \mathrm{e}^{2\mu + \sigma^2}$
$\mathrm{Median}[X] = \mathrm{e}^{\mu}$
$\mathrm{Mode}[X] = \mathrm{e}^{\mu - \sigma^2}$

Exponential distribution

$f_X(x; \lambda) = \lambda\mathrm{e}^{-\lambda x} \qquad \forall x \geq 0$

Properties of Exponential distribution

The Exponential distribution models the time between two events in a Poisson process with parameter λ.
$\mu_X = \frac{1}{\lambda}$ $\quad$ $\mathrm{Var}[X] = \frac{1}{\lambda^2}$

If $X_i \sim \text{Exponential}(\lambda)$ and they are independent
$\sum_i X_{i=1}^n \sim \Gamma(n, \lambda)$

Gamma distribution

$f_X(x; r, \lambda) = \frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1}\mathrm{e}^{-\lambda x} \qquad \forall x \geq 0$

Properties of Gamma distribution

The Gamma distribution models the time between $r+1$ events in a Poisson process with parameter λ.
Note that for $r = 1$ this is an Exponential distribution.
$\mu_X = \frac{r}{\lambda}$ $\quad$ $\mathrm{Var}[X] = \frac{r}{\lambda^2}$

χ^2 distribution

$f_X(x; k) = \frac{1}{\Gamma(\frac{k}{2})}(\frac{1}{2})^{\frac{k}{2}} x^{(\frac{k}{2}-1)}\mathrm{e}^{-\frac{x}{2}} \qquad \forall x \geq 0$

Properties of χ^2 distribution

The χ^2 distribution with k degrees of freedom (χ^2_k) models the sum of the standard distances between k normal samples and the respective means.
Note that $\chi^2_k \sim \text{Gamma}(\frac{k}{2}, \frac{1}{2})$
$\mu_X = k$ $\quad$ $\mathrm{Var}[X] = 2k$