# ST2132 Cheatsheet

for midterms, by ning

**Baye's Rule**
Suppose that $B_1, B_2, \cdots, B_n$ are partitions of the sample space $\Omega$. Then for any event $A$,

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{j=1}^n P(A|B_j)P(B_j)}$$

**Expectation**
The expectation of a random variable $X$ is defined as follows for the discrete and continuous case respectively,

$$\text{E}[X] = \sum_i x_i p(x_i)$$

$$\text{E}[X] = \int_{-\infty}^{\infty} x f(x)\ dx$$

**Moment Generating Functions**
The moment generating function (MGF) of a random variable $X$ is,

$$M(t) = \text{E}[e^{tX}]$$

and the $r^{\text{th}}$ moment of a random variable is $\text{E}[X^r]$ if it exists.

**Variance**
The variance $\sigma^2$ of a random variable $X$, then the variance of $X$ is,

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

And

$$Var(a + bX) = b^2 Var(X)$$

**Sample Variance**
The unbiased sample variance $S^2$ is

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

The biased sample variance $\hat{\sigma}^2$ is

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

**Covariance**
If $X$ and $Y$ are jointly distributed random variables with means $\mu_X$ and $\mu_Y$ respectively, then the covariance of $X$ and $T$ is,

$$Cov(X, Y) = \text{E}[(X - \mu_X)(Y - \mu_Y)]$$

If $X$ and $Y$ are independent, then

$$Cov(X, Y) = \text{E}[XY] - \text{E}[X]\,\text{E}[Y]$$

If $X$ and $Y$ are positively associated, then the covariance will be positive, and vice versa.

**Correlation**
Additionally, the correlation $\rho$ can be expressed as,

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

$-1 \le \rho \le 1$ and $\rho = \pm 1 \iff P(Y = a + bX) = 1$ for some constants $a, b$.

**Mean Square Error**
If the true value of a quantity being measured is denoted $x_0$, then the measurement $X$ can be modelled as,

$$X = x_o + \beta + \epsilon$$

where $\beta$ is the constant error and $\epsilon$ is the random component of the error. And

$$\text{E}[\epsilon] = 0$$
$$Var(\epsilon) = \sigma^2$$
$$\text{E}[X] = x_0 + \beta$$
$$Var(X) = \sigma^2$$

The mean squared error is then

$$MSE = \beta^2 + \sigma^2$$

**Bias and Standard Error**
The bias of an estimator is given by $\text{E}[\hat{\theta}] - \theta_0$. The standard error is the standard deviation of the sampling distribution.

**Bernoulli Distribution**
The Bernoulli distribution is defined over the parameter $p \in [0, 1]$. Its PMF is

$$P(X = x) = \begin{cases} 1 - p & \text{if } x = 0 \\ p & \text{if } x = 1 \end{cases}$$

The MGF is $1 - p + pe^t$. The mean and variance are $p$ and $p(1 - p)$ respectively. The fisher information is $1/(pq)$.

**Binomial Distribution**
The binomial distribution is defined over two parameters, $n \in \{0, 1, 2, \cdots\}$ and $p \in [0, 1]$. Its PMF is

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

The MGF is $(1 - p + pe^t)^n$. The mean and variance are $np$ and $np(1 - p)$ respectively. The fisher information is $\frac{n}{p(1-p)}$ for a fixed $n$.

**Poisson Distribution**
The poisson distribution is defined over the parameter $\lambda > 0$. Its PMF is

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

The MGF is $e^{\lambda(e^t - 1)}$. The mean and variance are both $\lambda$. The fisher information is $1/\lambda$.

**Geometric Distribution**
The geometric distribution is defined over the parameter $k \in \mathbb{Z}^+$. Its PMF is

$$P(X = k) = p(1 - p)^{k-1}$$

The MFG is $pe^t/(1 - (1 - p)e^t)$. The mean and variance are $1/p$ and $(1 - p)/p^2$ respectively.

**Gamma Distribution**
The gamma distribution is defined over two parameters, $\alpha > 0$, $\lambda > 0$. Its PDF is

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$$

The MGF is $\left(1 - \frac{t}{\lambda}\right)^{-\alpha}$ for $t < \lambda$. The mean and variance are $\alpha/\lambda$ and $\alpha/\lambda^2$ respectively.

**Normal Distribution**
The normal distribution is defined over two parameters, $-\infty < \mu < \infty$, $\sigma > 0$. Its probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty$$

The MGF is $e^{\mu t + \sigma^2 t^2/2}$. Its mean and variance are $\mu$ and $\sigma^2$ respectively. The normal distribution is symmetric about $\mu$, such that $f(\mu - x) = f(\mu + x)$.

**Standard Normal Distribution**
$Z \sim N(0, 1)$ is the standard normal. Its CDF is commonly denoted $\Phi$ and its density $\phi$. To 'standardise' a normal distribution $X$ to $Z$, note that

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

**$\chi^2$ Distribution**
For the standard random variable $Z$, the distribution of $Y = Z^2$ is called the chi-square distribution with 1 degree of freedom, $\chi_1^2$.

$\chi_1^2$ is a special case of the gamma distribution, where $\alpha = \lambda = 1/2$, i.e. $\chi_1^2 = \Gamma(1/2, 1/2)$.

Then, if $Y_1, Y_2, \cdots, Y_n$ are independent $\chi_1^2$ random variables, the distribution of $W = Y_1 + Y_2 + \cdots + Y_n$ is the $chi^2$ random variable with $n$ degrees of freedom, $\chi_n^2$.

The density of $\chi_n^2 \sim \Gamma(n/2, 1/2)$ is given by

$$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n/2)-1} e^{x/2}$$

**$t$ Distribution**
For the standard normal random variable $Z$ and $U \sim \chi_n^2$, where $Z$ and $U$ are independent, the distribution of $Z/\sqrt{U/n}$ is the $t$ distribution with $n$ degrees of freedom. Its PDF is

$$f(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\Gamma(n/2)}\left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}$$

**$F$ Distribution**
If $U \sim \chi_n^2$ and $V \sim \chi_m^2$, then the distribution of $W = \frac{U/n}{V/m}$ is the $F$ distribution with $n$ and $m$ degrees of freedom, $F_{n,m}$. Its PDF is

$$f(x) = \frac{\Gamma[(n+m)/2]}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{n}{m}\right)^{n/2} x^{n/2-1}\left(1 + \frac{n}{m}x\right)^{-(n+m)/2}$$

for $x \ge 0$. Also, if $T \sim t_n$ then $T^2 \sim F_{1,n}$

**Central Limit Theorem**
Let $X_1, X_2, \cdots$ be a sequence of independent random variables having mean 0 and variance $\sigma^2$ and the common distribution function $F$ and MGF $m$ defined in a neighbourhood of zero. If

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

then

$$\lim_{x\to\infty} P\left(\frac{S_n}{\sigma\sqrt{n}} \le x\right) = \Phi(x)$$

A more useful result is as follows: if $X_1, X_2, \cdots, X_n$ are i.i.d. random variables with large $n$, then

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

**Linear Functions of a Random Variable**
Let $Y = g(X)$. To find $f_Y(y)$,

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(g(X) \le y) \\ &= P(X \le g^{-1}(y)) \\ &= F_X(g^{-1}(y)) \\ f_Y(y) &= \frac{d}{dy} F_X(g^{-1}(y)) \\ &= \frac{dg^{-1}}{dy} f_X(g^{-1}(y)) \end{aligned}$$

**Non-linear Functions of Random Variables**
Let $Y = g(\vec{X})$, where $\vec{X} := (X_1, X_2, \cdots)$ with mean vector $\vec{\mu}$. Then, in order to find the mean and variance of $Y$, first take the Taylor expansion of $g(\vec{X})$,

$$\begin{aligned} Y &= g(\vec{X}) \\ &\approx g(\mu) + (X_1 - \mu_1)\frac{\partial g(\mu)}{\partial x_1} + (X_2 - \mu_2)\frac{\partial g(\mu)}{\partial x_2} + \cdots \end{aligned}$$

Then, $\text{E}[Y] \approx g(\mu)$, and

$$Var(Y) \approx Var(g(\mu) + (X_1 - \mu_1)\cdots$$

Consider for example, $\vec{X} := (X_1, X_2)$. Then

$$\begin{aligned} Var(X) \approx\ & \sigma_{X_1}^2 \left(\frac{\partial g(\mu)}{\partial x_1}\right)^2 + \\ & \sigma_{X_2}^2 \left(\frac{\partial g(\mu)}{\partial x_2}\right)^2 + \\ & 2\sigma_{XY}\left(\frac{\partial g(\mu)}{\partial x_1}\right)\left(\frac{\partial g(\mu)}{\partial x_2}\right) \end{aligned}$$

**Simple Random Sampling**
Simple random sampling *without replacement* means that each sample is not independent of another. While the mean of the simple random sample is still unbiased, that is $\text{E}[\bar{X}] = \mu$,

$$Cov(X_i, X_j) = -\sigma^2/(N - 1)$$

for two different simple random samples, i.e. $i \ne j$. The variance of the sample mean then becomes

$$Var(\bar{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

The variance of the sample total is

$$Var(T) = N^2\left(\frac{\sigma^2}{n}\right)\frac{N-n}{N-1}$$

For both expressions above, however, $\sigma$ is unknown and must be estimated. Therefore, we have also the unbiased estimates for $Var(\bar{X})$ and $Var(T)$

$$s_{\bar{X}}^2 = \frac{s^2}{n}\left(1-\frac{n}{N}\right)$$
$$s_T^2 = N^2 s_{\bar{X}}^2$$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is the unbiased sample variance.

**Method of Moments Estimators**
The method of moments estimates the parameter $\theta$ by finding expressions for it in terms of the lowest possible order moments and then substituting sample moments into these expressions.

**Maximum Likelihood Estimators**
The MLE estimator finds an estimate of the parameter $\theta_0$ which maximises the probability of having observed the sample. The likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

Often, this function is difficult to maximise. Since log is a monotonic increasing function, we may simplify this problem by finding the maximum of the loglikelihood function instead

$$l(\theta) = \sum_{i=1}^{n} \log f(x_i|\theta)$$

**Consistency**
Let $\hat{\theta}_n$ be an estimate of a parameter $\theta_0$ based on a sample of size $n$. $th\hat{e}ta_n$ is said to be consistent in probability if $\hat{\theta}_n$ converges in probability to $\theta_0$ as $n$ approaches infinity. That is, for $\epsilon > 0$,

$$P(|\hat{\theta}_n - \theta_0| > \epsilon) \to 0 \text{ as } n \to \infty$$

**Fisher Information**
The fisher information, $I(\theta)$ is defined as

$$I(\theta) = \mathrm{E}\left[\frac{\partial}{\partial\theta}\log f(X|\theta)\right]^2$$
$$= -\,\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right]$$

**Large Sample Theory for MLE**
Let $\hat{\theta}$ denote the MLE of $\theta_0$. The probability distribution of

$$\sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0)$$

tends to a standard normal distribution. Therefore, the asymptotic variance of the MLE is

$$\frac{1}{nI(\theta)} = -\frac{1}{\mathrm{E}[l''(\theta_0)]}$$

**Approximate Confidence Intervals**
Confidence intervals can be approximated through the large sample theory for MLE by taking $\sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0) \to N(0,1)$, as $n \to \infty$.

$$P\left(-z(\alpha/2) \leq \sqrt{nI(\hat{\theta})}(\hat{\theta} - \theta_0) \leq z(\alpha/2)\right) \approx 1 - \alpha$$