Probability and Statistics II formulae sheet

Probability

Distributions and related moment generating functions

1. $X \sim \mathrm{Binomial(n,p)}: \quad \mathrm{P(X=k)} = \binom{n}{k}\mathrm{p^k q^{n-k}},\ 0 \le \mathrm{k} \le \mathrm{n}$, where $q = 1-p$. $M_X(t) = (pe^t + q)^n$.
2. $X \sim \mathrm{Geometric(p)}: \quad \mathrm{P(X=k)} = \mathrm{qp^{k-1}},\ \mathrm{k} \ge 1$. $M_X(t) = \frac{p}{e^{-t}-q}$, where $q = 1-p$.
3. $X \sim \mathrm{Poisson}(\lambda): \quad \mathrm{P(X=k)} = \frac{\mathrm{e}^{-\lambda}\lambda^{\mathrm{k}}}{\mathrm{k!}},\ \mathrm{k} \ge 0$. $M_X(t) = e^{\lambda(e^t-1)}$.
4. $X \sim \mathrm{Ga}(\alpha, \beta)$, $\alpha > 0,\ \beta > 0$, if

$$f_X(x) = \begin{cases} \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} & \text{if } x > 0, \\ 0 & \text{if } x \le 0, \end{cases}$$

where $\Gamma(\alpha)$ is the Gamma function, which is given by $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$.

$M_X(t) = \frac{\beta^\alpha}{(\beta-t)^\alpha}$

Two particular cases of the Gamma distribution.

(a) $X \sim \mathrm{Exp}(\lambda)$ if $f_X(x) = \lambda e^{-\lambda x}$ if $x > 0$ and $f_X(x) = 0$ if $x \le 0$.

(b) $X \sim \chi^2(1)$ is equivalent to saying $X \sim \mathrm{Ga}(\frac{1}{2}, \frac{1}{2})$.

5. $X \sim \mathcal{N}(\mu, \sigma^2)$ if $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. $M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$.

Two more distributions

1. **Cauchy Distribution.** X has Cauchy distribution with *location* parameter x_0 and *scale* parameter γ and we write $X \sim Cauchy(x_0, \gamma)$ if

$$f_X(x) = \frac{1}{\pi\gamma}\left[\frac{\gamma^2}{(x-x_0)^2 + \gamma^2}\right],$$

2. **(Student's) t Distribution.** X has t distribution with ν degrees of freedom, and we write $X \sim t_\nu$, if

$$f_X(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

Expectations

1. If X is a random variable, $g : \mathbb{R} \mapsto \mathbb{R}$ is a function, and $Y = g(X)$ is a new random variable, then $\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\, ds$.
2. Similarly, if $Z = g(X,Y)$, where (X,Y) are random variables with pdf $f_{X,Y}(x,y)$ then

$$\mathbb{E}(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\, dxdy.$$

3. If $f_{X|Y=y}(x)$ is the pdf of X conditioned on $Y = y$, then $\mathbb{E}(g(X)|Y=y) = \int_{-\infty}^{\infty} g(x) f_{X|Y=y}(x)\, dx$.

Transformation of random variables

1. If $Y = g(X)$, then $f_Y(y) = f_X(g^{-1}(y))|\frac{dg^{-1}(y)}{dy}|$
2. If $U = g_1(X,Y),\ V = g_2(X,Y)$ and the inverse transformation is given by $X = h_1(U,V),\ Y = h_2(U,V)$, then

$$f_{U,V}(u,v) = \begin{cases} f_{X,Y}(h_1(u,v), h_2(u,v))\,|J|, & \text{for } (u,v) \in B, \\ 0, & \text{otherwise,} \end{cases}$$

where B is the range of (U,V), $J = \begin{vmatrix} \frac{\partial h_1}{\partial u} & \frac{\partial h_2}{\partial u} \\ \frac{\partial h_1}{\partial v} & \frac{\partial h_2}{\partial v} \end{vmatrix}$.

Statistics

Sample mean

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Sample variance as

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

T-distribution: let $W \sim N(0,1)$ and $V \sim \chi_r^2$ be independent

$$T = \frac{W}{\sqrt{V/r}} \sim t_r$$

Testing the mean when the variance is known

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

Test of hypothesis and confidence interval for the variance

$$W = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi_{n-1}^2$$

Test for a normal mean when variance is unknown

$$T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S} \sim t_{n-1}$$

Hypothesis tests for a Poisson mean

$$T = \frac{\bar{X} - \lambda_0}{\sqrt{\lambda_0/n}} \sim N(0,1)$$

Goodness of fit tests

$$X^2 = \sum_{i=1}^{k}\frac{(O_i - E_i)^2}{E_i}$$

Test for the mean of two independent samples when the variance is known

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma\sqrt{1/n_1 + 1/n_2}} \sim N(0,1)$$

Test for the mean of two independent samples when the variance is unknown

$$T = \frac{\bar{X} - \bar{Y}}{S_0\sqrt{1/n_1 + 1/n_2}} \sim t_{n_1+n_2-2}$$

F test for comparing two variances

$$F = \frac{S_1^2}{S_2^2} \sim F_{n_2-1}^{n_1-1}$$

An approximate test when variances are unequal

$$T^* = \frac{\bar{X} - \bar{Y}}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \approx t_{\frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\left(\frac{s_1^4/n_1^2}{n_1-1} + \frac{s_2^4/n_2^2}{n_2-1}\right)}}$$

Matched pairs t-test

$$T = \frac{\bar{d}\sqrt{n}}{s_d} \sim t_{n-1}$$

Test of two proportions

$$Z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

Comparing two correlation coefficients

$$Z' = \frac{1}{2}\ln\left[\frac{1+r}{1-r}\right] \approx \text{Normal}(\frac{1}{2}\ln\left[\frac{1+\rho}{1-\rho}\right], 1/(n-3))$$