ST2334 Cheat Sheet

Marginal Distributions

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

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy$$

Conditional Distributions

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

Independent Random Variables

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \text{ and } f_{X|Y}(x \mid y) = f_X(x)$$

Expectation

$$E[g(X, Y)] = \sum_x \sum_y g(x, y)f_{X,Y}(x, y)$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f_{X,Y}(x, y)\, dy\, dx$$

Covariance

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

1. $Cov(X, Y) = E(XY) - E(X)E(Y) = E(XY) - \mu_X \mu_Y$
2. $Cov(X, X) = Var(X)$
3. $Cov(X, Y) = Cov(Y, X)$
4. $Cov(aX + b, cY + d) = ac\, Cov(X, Y)$
5. $Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2ab\, Cov(X, Y)$

If X and Y are independent, then their covariance is 0.

Correlation Coefficient

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

1. $-1 \leq \rho_{X,Y} \leq 1$
2. It is the measure of degree of linear relationship.
3. If X and Y are independent, then $\rho_{X,Y} = 0$.

Discrete Uniform Distribution

$$f_X(x) = P(X = x) = \begin{cases} \frac{1}{k} & x = x_1, x_2, \cdots, x_k \\ 0 & \text{otherwise} \end{cases}$$

$$\mu = \frac{1}{k}\sum_{i=1}^{k} x_i,\ \sigma^2 = \frac{1}{k}\sum_{i=1}^{k}(x_i - \mu)^2 = \frac{1}{k}\sum_{i=1}^{k} x_i^2 - \mu^2$$

Bernoulli Distribution

$$f_X(x) = P(X = x) = p^x(1 - p)^{1-x}, \text{ for } x = 0, 1$$
$$\mu = p, \sigma^2 = p(1 - p)$$

Binomial Distribution

$$f_X(x) = P(X = x) = \binom{n}{x} p^x(1 - p)^{n-x}, \text{ for } x = 0, 1, \cdots, n$$
$$\mu = np, \sigma^2 = np(1 - p)$$

When $n = 1$, Binomial distribution is Bernoulli distribution.

Geometric Distribution

$$X \sim Geom(p)$$
$$f_X(x) = P(X = x) = (1 - p)^{x-1}p, \text{ for } x = 1, 2, \cdots$$
$$\mu = \frac{1}{p}, \sigma^2 = \frac{1 - p}{p^2},\ F(x) = 1 - (1 - p)^x$$

$$P(X > n + k \mid X > n) = P(X > k)$$

Negative Binomial Distribution

$$X \sim NB(k, p)$$
$$f_X(x) = P(X = x) = \binom{x - 1}{k - 1} p^k q^{x-k}, \text{ for } x = k, k + 1, \cdots$$
$$\mu = \frac{k}{p}, \sigma^2 = \frac{(1 - p)k}{p^2}$$

Note that $Geom(p) = NB(1, p)$

Poisson Distribution

$$f_X(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$
$$\mu = \lambda, \sigma^2 = \lambda$$

Poisson Approximation of Binomial

Let $X \sim B(n, p)$. If $n \to \infty$ and $p \to 0$ such that $\lambda = np$ remains a constant, then $X \sim Poisson(np)$.
($n \geq 20$ and $p \leq 0.05$) or ($n \geq 100$ and $np \leq 10$)

Continuous Uniform Distribution

$$X \sim U(a, b)$$
$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$
$$\mu = \frac{a + b}{2}, \sigma^2 = \frac{(b - a)^2}{12}$$

Exponential Distribution

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$
$$\mu = \frac{1}{\lambda}, \sigma^2 = \frac{1}{\lambda^2}$$
$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$
$$P(X > s + t \mid X > s) = P(X > t)$$

Normal Distribution

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

Standard Normal

$$\phi = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \Phi = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy$$

$$\text{If } Y \sim N(\mu, \sigma^2), P(a < Y \leq b) = \Phi(\frac{b - \mu}{\sigma}) - \Phi(\frac{a - \mu}{\sigma})$$

1. $P(Z \geq 0) = P(Z \leq 0) = 0.5$
2. $-Z \sim N(0, 1)$
3. $P(Z \leq x) = 1 - p(Z > x)$
4. $P(Z \leq -x) = P(Z \geq x)$
5. If $Y \sim N(\mu, \sigma^2)$, then $X = \frac{Y-\mu}{\sigma} \sim N(0, 1)$.
6. If $X \sim N(0, 1)$, then $Y = aX + b \sim N(b, a^2)$.

Normal Approximation to Binomial

If $np > 5$ and $n(1 - p) > 5$, and X is a binomial random variable with $\mu = np$ and $\sigma^2 = np(1 - p)$, then as $n \to \infty$,

$$Z = \frac{X - np}{\sqrt{np(1 - p)}} \sim N(0, 1)$$

Continuity Correction

1. $P(X = k) \approx P(k - 0.5 < X < k + 0.5)$
2. $P(X \geq k) \approx P(X \geq k - 0.5)$
3. $P(X \leq k) \approx P(X \leq k + 0.5)$

Sampling

With replacement, P(each sample selected) $= \frac{1}{N^n}$. Without replacement, P(each sample selected) $= \frac{1}{\binom{N}{n}}$.

Generally, $\mu_{\overline{X}} = \mu_X$ and $\sigma^2_{\overline{X}} = \frac{\sigma^2_X}{n}$.
LLN $P(\mid \overline{X} - \mu \mid > \epsilon) \to 0$ as $n \to \infty$.
CLT $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ or $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$
If $X_i, i = 1, 2, \cdots, n$ are $N(\mu, \sigma^2)$, then $\overline{X}$ is $N(\mu, \sigma^2/n)$ regardless of sample size.

Different Samples

If two samples are independent,

$$E(\overline{X} - \overline{Y}) = \mu_{\overline{X}-\overline{Y}} = \mu_{\overline{X}} - \mu_{\overline{Y}}$$
$$V(\overline{X} - \overline{Y}) = \sigma^2_{\overline{X}-\overline{Y}} = \frac{\sigma_1^2}{n_{\overline{X}}} + \frac{\sigma_2^2}{n_{\overline{Y}}}$$

Gamma function

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1}e^{-y}\, dy$$

$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1), \Gamma(1) = 1, \Gamma(n) = (n - 1)!$ where n is an integer.

χ^2-Distribution

$$f_Y(y) = \frac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1}e^{-y/2}, y > 0$$

1. If $Y \sim \chi^2(n)$, then $E(Y) = 2$ and $V(Y) = 2n$
2. For large n, $\chi^2(n) \sim N(n, 2n)$
3. If $Y_1, \cdots, Y_k$ are independent chi-square random variables, then $Y_1 + \cdots + Y_k$ is a chi-square distribution with $n_1 + \cdots + n_k$ degrees of freedom.
4. If $Z \sim N(0, 1)$, then $Z^2 \sim \chi^2(1)$
5. If $X \sim N(\mu, \sigma^2)$, then $(\frac{X-\mu}{\sigma})^2 \sim \chi^2(1)$
6. If $X_1, \cdots, X_n$ are random sample from a normal population with mean μ and variance σ^2, then $Y = \sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2 \sim \chi^2(n)$

Sample Variance

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

Let S^2 be the sample variance of a random sample of size n taken from a normal population with mean μ and variance σ^2. Then,

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

t-Distribution

Let Z be a standard normal variable and U a χ^2 random variable with n degrees of freedom. If Z and U are independent, then T is $T = \frac{Z}{\sqrt{U/n}}$.

$$f_T(t) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)}(1 + t^2/n)^{(n+1)/2}$$

1. If $T \sim t(n)$, then $E(T) = 0$ and $V(T) = \frac{n}{n-2}$ for $n > 2$.
2. t-distribution approaches $N(0, 1)$ as $n \to \infty$. Replace when $n \geq 30$.
3. $T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

F-Distribution

Let U and V be independent random variables having χ^2 distributions with n_1 and n_2 degrees of freedom. Then, $F = \frac{U/n_1}{V/n_2}$ is called a F distribution with (n_1, n_2) degrees of freedom.

$$f_F(x) = \frac{n_1^{n_1/2}n_2^{n_2/2}\Gamma((n_1+n_2)/2)x^{n_1/2-1}}{\Gamma(n_1/2)\Gamma(n_2/2)(n_1x+n_2)^{(n_1+n_2)/2}}, x > 0$$

1. If $F \sim F(n, m)$, then $1/F \sim F(m, n)$
2. $F(n_1, n_2; \alpha) = 1/F(n_2, n_1; 1-\alpha)$

CI for μ with known σ

If $\overline{X}$ is the mean of a random sample of size n from a population with known variane σ^2, a $(1-\alpha)100\%$ confidence internval for μ is given by

$$P(Z > z_\alpha) = \alpha$$

$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = (\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

$$P(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

For a given margin of error e, the sample size is given by $n \geq (\frac{z_{\alpha/2}}{e})^2\sigma^2$

CI for μ with unknown σ, normal population

If $n < 30$,

$$\overline{X} \pm t_{n-1;\alpha/2}\frac{S}{\sqrt{n}} = (\overline{X} - t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}\overline{X} + t_{n-1;\alpha/2}\frac{S}{\sqrt{n}})$$

If $n > 30$,

$$\overline{X} \pm z_{\alpha/2}\frac{S}{\sqrt{n}} = (\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}})$$

CI for $\mu_1 - \mu_2$ with known or unknown $\sigma_1^2 \neq \sigma_2^2$

Either the two populations are normal or sample sizes are more than 30.

$$(\overline{X_1} - \overline{X_2}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

For unknown σ^2 and big n, replace σ^2 with their estimates S^2.

CI for $\mu_1 - \mu_2$ with unknown $\sigma_1^2 = \sigma_2^2$

For small n

$$(\overline{X_1} - \overline{X_2}) \pm t_{n_1+n_2-2;\alpha/2}S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

For large n

$$(\overline{X_1} - \overline{X_2}) \pm z_{\alpha/2}S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

CI for $\mu_D = \mu_1 - \mu_2$, normal population

For small n,

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i), \overline{D} \pm t_{n-1;\alpha/2}\frac{S_D}{\sqrt{n}}$$

For large n,

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i), \overline{D} \pm z_{\alpha/2}\frac{S_D}{\sqrt{n}}$$

CI for σ^2, normal population

If μ is known,

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n;\alpha/2}} < \sigma^2 < \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n;1-\alpha/2}}$$

If μ is unknown,

$$\frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}$$

CI for σ_1^2/σ_2^2, normal population, unknown μ

$$\frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1,n_2-1;\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}F_{n_2-1,n_1-1;\alpha/2}$$

Error types

The rejection of H_0 when it's true is called type I error and its probability is level of significance α.
Not rejecting H_0 when it's false is called type II error and its probability is β and the power of the test is $1 - \beta$.

Test Statistic; known or unknown σ; Normal population or large n

When H_0 is true,

$$Z = \frac{\overline{X_1} - \overline{X_2} - \delta_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1)$$

Replace σ^2 with S^2 for unknown σ.

Test Statistic; unknown $\sigma_1^2 = \sigma_2^2$; Normal population and small n

When H_0 is true,

$$T = \frac{(\overline{X_1} - \overline{X_2}) - \delta_0}{S_p\sqrt{1/n_1 + 1/n_2}} \sim t(n_1 + n_2 - 2)$$

Paired Data

$$D_i = X_i - Y_i, \mu_D = \mu_1 - \mu_2$$

$$T = \frac{\overline{D} - \mu_{D,0}}{S_D/\sqrt{n}} \sim t(n-1), n < 30$$

$$T = \frac{\overline{D} - \mu_{D,0}}{S_D/\sqrt{n}} \sim N(0, 1), n >= 30$$

Hypothesis Test on σ^2

To test $H_0 : \sigma^2 = \sigma_0^2$,

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

Hypothesis Test on σ_1^2, σ_2^2

To test $H_0 : \sigma_1^2 = \sigma_2^2$,

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

Chebyshev & Tail-Sum

$$P(\mid X - \mu \mid > k\sigma) \leq \frac{1}{k^2}$$

$$P(\mid X - \mu \mid \leq k\sigma) \geq 1 - \frac{1}{k^2}$$

$$E(M) = \sum_{k=1}^{\infty} P(M \geq k)$$

Compiled by **Mohideen Imran Khan**.
Template from http://wch.github.io/latexsheet/