MATH 321 COLLECTED FORMULAS

A. Probability

Method. The number of ways to select k elements from an n-element set is...

	Order matters	Order doesn't matter
With replacement	n^k	$\binom{n+k-1}{k}$
Without replacement	$\frac{n!}{(n-k)!}$	$\binom{n}{k} = \frac{n!}{(n-k)!k!}$

Theorem. *Properties of (all) probabilities:*

(1) $P(\emptyset) = 0$
(2) $P(A) = 1 - P\left(A^C\right)$
(3) *If* $A \subseteq B$, *then* $P(A) \leq P(B)$
(4) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Definition. Let A and B be events with $P(B) \neq 0$. The **conditional probability of** A **given** B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Definition. Events A and B are **independent** if and only if $P(A \cap B) = P(A)P(B)$.

Theorem (Multiplication rule for probabilities)**.** *Let* A *and* B *be events with* $P(B) \neq 0$. *Then*

$$P(A \cap B) = P(A|B)P(B)$$

Theorem (The Law of Total Probability)**.** *If event* B *has probability strictly between* 0 *and* 1, *then*

$$P(A) = P(A|B)P(B) + P(A|B^C)P(B^C)$$

Theorem (Bayes' Law)**.** *If* A *and* B *are events with positive probability, then*

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

B. Random Variables

Definition. A **random variable** X assigns a number to each outcome in the sample space S.

(1) All random variables have a **cumulative distribution function (CDF)**: $F(x) = P(X \leq x)$.
(2) A discrete random variable has a **probability mass function (PMF)**: $p(x) = P(X = x)$.
(3) A continuous random variable has a **probability density function (PDF)** $f(x)$ such that for any numbers a and b (with $a \leq b$)

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Definition. Expected value (or **mean**):

(1) If X is a discrete RV with PMF $p(x)$, then $\mu = E(X) = \sum_x xp(x)$.
(2) If X is a continuous RV with PDF $f(x)$, then $\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$.

Definition. Variance: $\sigma^2 = \mathrm{Var}(X) = E\left[(X-\mu)^2\right] = E(X^2) - [E(X)]^2$. **Standard deviation**: $\sigma = \sqrt{\sigma^2}$.

Theorem. *For any random variable* X *and any constants* a *and* b*:*

(1) $E(aX + b) = aE(X) + b$ *and*
(2) $Var(aX + b) = a^2\, Var(X)$.

Theorem. *If* $X_1, X_2, \ldots X_n$ *are independent, then*

(1) $E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$
(2) $Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots + Var(X_n)$

Date: December 11, 2019.

C. Statistics

C.1. Sampling.

Definition. A **random sample of size** n is a set of independent identically distributed random variables $X_1, X_2, \ldots X_n$. Some **sample statistics**:

(1) The **sample mean**: $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$

(2) The **sample variance**: $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

Theorem. *For any random sample from a population with mean μ and variance σ^2:*
(1) $E(\overline{X}) = \mu$ and $Var(\overline{X}) = \frac{\sigma^2}{n}$
(2) $E(S^2) = \sigma^2$

Definition. The **sample standard error** is $\dfrac{s}{\sqrt{n}}$

Definition. A sample statistic $\hat{X}$ is an **unbiased estimator** of population parameter ρ if $E(\hat{X}) = \rho$.

Theorem. *If $\overline{X}$ is a the mean of a random sample from a normally distribution population, then $\overline{X}$ is normally distributed (with mean and variance given in the last theorem).*

Theorem (Central Limit Theorem). *If $\overline{X}$ is a the mean of a random sample from a population, then $\overline{X}$ is approximately normally distributed (with mean and variance given in the theorem above).*

C.2. Confidence (and prediction) intervals.

1. $100(1-\alpha)\%$ CI for μ (known σ): $\overline{x} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$
2. $100(1-\alpha)\%$ CI for μ (large sample): $\overline{x} \pm z_{\alpha/2}\dfrac{s}{\sqrt{n}}$
3. $100(1-\alpha)\%$ CI for μ (normal population): $\overline{x} \pm t_{\alpha/2,n-1}\dfrac{s}{\sqrt{n}}$
4. $100(1-\alpha)\%$ **prediction interval** for μ (normal population): $\overline{x} \pm t_{\alpha/2,n-1}\sqrt{\dfrac{s^2(n+1)}{n}}$
5. $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ (known σ_1 and σ_2, normal populations): $\overline{x} - \overline{y} \pm z_{\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$
6. $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ (large samples): $\overline{x} - \overline{y} \pm z_{\alpha/2}\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$
7. $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ (normal populations with the same variance): $\overline{x} - \overline{y} \pm t_{\alpha/2,n_1+n_2-2}\sqrt{s_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$

   $s_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$ is the pooled estimator of the common variance
8. $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ (normal populations with difference variances): $\overline{x} - \overline{y} \pm t_{\alpha/2,\nu}\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

   $\nu \approx \dfrac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$ (round down to the nearest integer)
9. Approximate $100(1-\alpha)\%$ CI for a proportion θ (large sample; x and $n-x$ both large):
   $\dfrac{x}{n} \pm z_{\alpha/2}\sqrt{\dfrac{1}{n}\left(\dfrac{x}{n}\right)\left(1-\dfrac{x}{n}\right)}$

C.3. **Test Statistics.**

For **tests about the mean** ($H_0 : \mu = \mu_0$) test statistics are:

- $z = \dfrac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ (known variance σ^2, all sample sizes if the pop. is normal, otherwise just large samples)
- $t = \dfrac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ (samples from approximately normally distributed populations, $n-1$ degrees of freedom)

  R command: `t.test(x)`

For **tests about the difference of two means** ($H_0 : \mu_1 - \mu_2 = \delta_0$) some test statistics are:

- $z = \dfrac{\overline{x}_1 - \overline{x}_2 - \delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ (known variances, all sample sizes if pops are normal, otherwise just large samples)
- $t = \dfrac{\overline{x}_1 - \overline{x}_2 - \delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ (normally populations with the same variance, $n_1 + n_2 - 2$ d.f.).

  $$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

  R command: `t.test(x, y, var.equal = T)`
- $t = \dfrac{\overline{x}_1 - \overline{x}_2 - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ (normally populations with different variances, ν degrees of freedom)

  $$\nu \approx \frac{\left(\frac{s_1^2}{n_2} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

  R command: `t.test(x, y)`

For **tests about a population proportion** ($H_0 : \theta = \theta_0$) we can use the sample proportion $\hat{\Theta} = X/n$ or the sample total $X = n\hat{\Theta}$ and the test statistics are:

- x (X is binomial with parameters n and θ_0)

  R command : `binom.test(x, n, p=`θ_0`)`
- $z = \dfrac{\hat{\theta} - \theta_0}{\sqrt{\frac{1}{n}\theta_0(1-\theta_0)}} = \dfrac{x - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}}$ (large samples, both $n\theta_0 \geq 10$ and $n(1-\theta_0) \geq 10$)

For **tests about the variance** ($H_0 : \sigma^2 = \sigma_0^2$) the test statistic is :

- $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$ (chi-square distribution, $n-1$ degrees of freedom)

For **tests about the ratio of two variances** ($H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$) the test statistic is

- $\dfrac{s_1^2}{s_2^2}$ (F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom, order matters).

  R command: `var.test(x, y)`

C.4. **Linear regression.**

Model (Linear regression). $\mu_{Y|X=x} = \alpha_1 + \beta_1 x$ or $\boxed{y = \alpha_1 + \beta_1 x + \epsilon}$. For most regression analysis we require $\epsilon \sim N(0, \sigma_\epsilon^2)$.

Verify that the linear model is reasonable by looking at a plot of your data: `> plot(y~x)`.

C.4.1. *Regression statistics.* Sample: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. Many of the statistics are calculated by R: `> model<-lm(y~x)` and `> summary(model)` will be useful.

$$\overline{x} = \frac{1}{n}\sum x_i \qquad \overline{y} = \frac{1}{n}\sum y_i$$

$$\hat{\alpha}_1 = \overline{y} - \hat{\beta}_1\overline{x} \qquad \hat{\beta}_1 = \frac{S_{XY}}{S_{XX}}$$

$$S_{XX} = \sum(x_i - \overline{x})^2 \qquad S_{XY} = \sum(x_i - \overline{x})(y_i - \overline{y})$$

$$\hat{y}_i = \hat{\alpha}_1 + \hat{\beta}_1 x_i \qquad i^{\text{th}} \text{ residual: } e_i = y_i - \hat{y}_i$$

$$SST = S_{YY} = \sum(y_i - \overline{y})^2 \qquad SSE = \sum\left[y_i - (\hat{\alpha}_1 + \hat{\beta}_1 x_i)\right]^2 = \sum e_i^2$$

$$SSR = SST - SSE = \sum(\hat{y}_i - \overline{y})^2 \qquad s_\epsilon^2 = \frac{SSE}{n-2} \text{ (note: } s_\epsilon \text{ is residual standard error)}$$

$$\text{Coefficient of determination } r^2 = \frac{SSR}{SST} \qquad \text{Sample correlation } r = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}} = \pm\sqrt{r^2}$$

C.4.2. *Test statistics and confidence intervals.* All assume $\epsilon \sim N(0, \sigma_\epsilon^2)$; you should check on this assumption before proceeding using the plots of residuals vs fitted values and normal Q-Q: `> plot(model)`.

Test and interval concerning β_1. Hypothesis test $H_0 : \beta_1 = c$. Test stat: $t = \dfrac{\hat{\beta}_1 - c}{\frac{s_\epsilon}{\sqrt{S_{XX}}}}$ has a t dist with $n-2$ df. R tests $H_0 : \beta_1 = 0$ against $H_0 : \beta_1 \neq 0$ by default. $100(1-\alpha)\%$ confidence interval for β_1:

$$\hat{\beta}_1 \pm t_{1-\alpha/2,n-2}\frac{s_\epsilon}{\sqrt{S_{XX}}}$$

CI and PI for the regression line. $100(1-\alpha)\%$ confidence interval for $\mu_{Y|X=x}$:

$$(\hat{\alpha}_1 + \hat{\beta}_1 x) \pm (t_{1-\alpha/2,n-2})(s_\epsilon)\sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{S_{XX}}}$$

$100(1-\alpha)\%$ prediction interval for Y given $X = x$:

$$(\hat{\alpha}_1 + \hat{\beta}_1 x) \pm (t_{1-\alpha/2,n-2})(s_\epsilon)\sqrt{1 + \frac{1}{n} + \frac{(x-\overline{x})^2}{S_{XX}}}$$

D. Special Discrete Distributions

Definition. A random variable X has a **discrete uniform distribution** if it is equally likely to assume any one of a finite set of possible values.

Definition. A random variable X has a **Bernoulli distribution** with parameter θ (with $0 < \theta < 1$) if its probability mass function is

$$\boxed{m(x) = \begin{cases} 1-\theta & \text{if } x = 0 \\ \theta & \text{if } x = 1 \end{cases}}$$

The outcome 1 is often referred to as "success" while 0 is "failure" and the experiment is often called a Bernoulli trial.

Proposition. The mean and variance of a Bernoulli random variable are $\mu = \theta$ and $\text{Var} = \theta(1-\theta)$.

Definition. The total number of successes in n independent, identically distributed (iid) Bernoulli trials is a random variable with a **Binomial distribution**. The probability mass function of a random variable X having a binomial distribution with parameters n and θ is

$$\boxed{b(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \text{ for } x = 0, 1, \ldots, n}$$

Proposition. The mean and variance of a binomial distribution are $\mu = n\theta$ and $\text{Var} = n\theta(1-\theta)$.

Definition. Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed (iid) Bernoulli trials, all with probability of success θ. Let N be the trial on which the first success occurs. The random variable N is said to have a **geometric distribution** with parameter θ and its probability mass function is

$$\boxed{g(n) = \theta(1-\theta)^{n-1} \text{ for } n = 1, 2, 3, \ldots}$$

Proposition. The mean and variance of a geometric distribution are $\mu = \dfrac{1}{\theta}$ and $\text{Var} = \dfrac{1}{\theta}\left(\dfrac{1}{\theta} - 1\right)$.

Definition. A random variable with the probability mass function

$$\boxed{p(x) = \frac{\lambda^x e^{-\lambda}}{x!} \text{ for } x = 0, 1, 2, \ldots}$$

is said to have a **Poisson distribution** with parameter $\lambda > 0$.

Proposition. The mean and variance of a Poisson distribution are $\mu = \lambda$ and $\text{Var} = \lambda$.

E. Special Continuous Distributions

Definition. A random variable X with a **uniform continuous distribution** with parameters α and β (with $\alpha < \beta$) has the following probability density function: $\boxed{f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{if } \alpha < x < \beta \\ 0 & \text{elsewhere} \end{cases}}$.

Proposition. A uniform continuous distribution with parameters α and β has mean $\mu = \frac{\alpha+\beta}{2}$, and variance $\sigma^2 = \frac{(\beta-\alpha)^2}{12}$.

Definition. A random variable with an **exponential distribution** with parameter $\theta > 0$ has the following probability density function: $\boxed{g(x) = \begin{cases} \frac{1}{\theta}e^{-x/\theta} & \text{if } x > 0 \\ 0 & \text{elsewhere} \end{cases}}$

Proposition. An exponential distribution with parameter θ has mean $\mu = \theta$ and variance $\sigma^2 = \theta^2$.

Definition. A random variable with a **normal distribution** with parameters μ and $\sigma > 0$ has the following probability density function: $\boxed{n(x; \mu, \sigma) = \dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ for all } x \in \mathbb{R}}$.

Proposition. A normal distribution with parameters μ and σ has mean $\mu = \mu$ and variance $\sigma^2 = \sigma^2$.