Summary of Equations

Set Basics

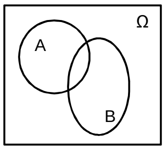
$A \cap B = \{x | x \in A \; and \; x \in B\}$ $A \cup B = \{x | x \in A \; or \; x \in B\}$

Useful Identities: $\Omega^c = \phi$ $A \cup A^c = \Omega$

$(A^c)^c = A$ $A \cap A^c = \phi$

DeMorgan's Laws: $\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i{}^c$

$\left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i{}^c$



Summary of Equations

Probability Axioms and Identities

- Axioms of Probability
  - Non-negativity: $\mathbb{P}(E) \geq 0$ for all E
  - Normalization: $\mathbb{P}(\Omega) = 1$
  - Additivity: for disjoint events, $\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \cdots + \mathbb{P}(E_n)$
- Given any probability law that obeys the probability axioms,
  - $\mathbb{P}(\emptyset) = 0$
  - $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$
  - $\mathbb{P}(E) \leq 1$
  - If $E \subset F$ then $\mathbb{P}(E) \leq \mathbb{P}(F)$
  - $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$

Summary of Equations

Combinatorics

- Draw k from a set of n with replacement
  - Order matters: simple product rule = n^k
  - Order doesn't matter: $n^k/k!$
- Draw k from a set of n without replacement
  - Order matters: k-permutations ${}_nP_k = n!/(n-k)!$
  - Order doesn't matter: $\binom{n}{k} = \frac{{}_nP_k}{k!} = \frac{n!}{k!\,(n-k)!}$
  - General problem: break it into stages where each stage is one of the above types

Summary of Equations

Probability

- Conditional Probability: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$
- Total Probability:
  - Let $A_1 \ldots A_n$ be a partition of the sample space

$$\mathbb{P}(B) = \sum \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$

Example: $\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)$

- Bayes' Rule: $\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$

Summary of Equations

Random Variables PMF and PDF

Discrete RV

$$\mathbb{P}(a \leq X \leq b) = \sum_{a \leq x \leq b} p_X(x)$$

$$\sum_{all\,k} p_X(k) = 1$$

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{all\,k \leq x} p_X(k)$$

Continuous RV

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx$$

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

$$F_X(b) = \int_{-\infty}^{b} f_X(x)dx$$

$$f_X(x) = \frac{dF_X(x)}{dx}$$

Summary of Equations

Expectation and Variance

Discrete:

$$E[X] = \sum_{all\,x} x\, p_X(x) \qquad var[X] = \sum_{all\,x} (x - E[X])^2\, p_X(x)$$

Continuous:

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)dx \qquad var[X] = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x)dx$$

Bernoulli Random Variable

- Consider a coin toss that produces a head (success) with probability *p*

$$X = \begin{cases} 1 & if\ heads\ (success) \\ 0 & if\ tails\ (failure) \end{cases} \qquad p_X(k) = \begin{cases} p & if\ k=1 \\ 1-p & if\ k=0 \end{cases}$$

$$E[X] = \sum_{all\ x} x\, p_X(x) = p$$

$$var[X] = \sum_{all\ x} (x - E[X])^2\, p_X(x) = p(1-p)$$

Binomial Distribution

- Repeat a Bernoulli trial *n* times (*p* = probability of success), X = number of successes

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

$$E[X] = n\,p$$

$$var[X] = n\,p(1-p)$$

Estimator for p

$$\hat{p} = \frac{X}{n} \qquad E(\hat{p}) = p$$

$$var(\hat{p}) = \frac{pq}{n}$$

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Geometric Distribution

- How many coin tosses are required before the first heads (success) comes up?
- Let X = number of tosses to get the first head

$$p_X(k) = (1-p)^{k-1} p$$

$$E[X] = 1/p$$

$$var[X] = (1-p)/p^2$$

Poisson Distribution

- Consider the Binomial distribution when n >> k
- More specifically, let *n* → ∞ while keeping *np* = λ = constant

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

$$E[X] = \lambda$$

$$var[X] = \lambda$$

Continuous Uniform PDF

- Uniform pdf between a and b, zero outside this range

$$f_X(x) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & otherwise \end{cases}$$

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \frac{b+a}{2}$$

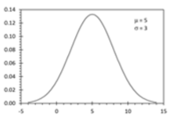
$$var[X] = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) dx = \frac{(b-a)^2}{12}$$

Normal Distribution

- Also called the Gaussian distribution

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

$$F_X(x) = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right]$$



$$E[X] = \mu$$

$$var[X] = \sigma^2$$

Summary of Equations

Sampling Distribution of the Mean

$$\bar{X} = \frac{1}{n}\sum_{i=1,n} X_i \qquad E[\bar{X}] = \mu \qquad var[\bar{X}] = \frac{\sigma^2}{n}$$

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad \text{or} \quad Student's\ t = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

Confidence Interval:

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}} \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

 Online Review Course of Undergraduate Probability and Statistics 13

Summary of Equations

Independent and Pooled Samples

- Two independent samples ($\bar{X}_1$, S_1) and ($\bar{X}_2$, S_2):

$$E[\overline{X_1} - \overline{X_2}] = \mu_1 - \mu_2 \qquad var[\overline{X_1} - \overline{X_2}] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

- Pooled Samples:

$$var[\overline{X_1} - \overline{X_2}] = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

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

 Online Review Course of Undergraduate Probability and Statistics 14

Summary of Equations

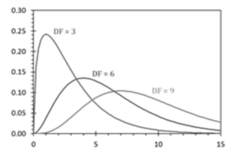
Sampling Distribution of the Variance

- For $X_i \sim N(\mu, \sigma^2)$ $\qquad \chi^2 = \frac{(n-1)S^2}{\sigma^2}$

$$E[\chi^2] = n - 1 \qquad var[\chi^2] = 2(n-1)$$

Confidence Interval:

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



 Online Review Course of Undergraduate Probability and Statistics 15

Summary of Equations

Linear Regression

- Model: $E[Y|X] = aX + b,\ \ y_i = ax_i + b + \epsilon_i$

$$r = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}} \qquad R^2 = r^2 = 1 - \frac{var(\epsilon)}{var(Y)}$$

$$a = \frac{cov(X,Y)}{var(X)} \qquad b = E[Y] - aE[X]$$

$$cov(x,y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

 Online Review Course of Undergraduate Probability and Statistics 16

Summary of Equations

Linear Regression Estimators

$$r = \frac{1}{n-1}\sum_{i=1}^{n} z_{xi}z_{yi} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{(n-1)s_x s_y}$$

$$cov(x,y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{(n-1)}$$

$$a = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \qquad b = \bar{y} - a\bar{x}$$

$$SSE = \sum_{i=1}^{n}\epsilon_i^2 \qquad s_\epsilon^2 = \frac{SSE}{n-2}$$

 Online Review Course of Undergraduate Probability and Statistics 17

Summary of Equations

Uncertainty of Linear Regression Fit

For $\epsilon_i \sim N(0, \sigma_\epsilon^2)$

$$\hat{y}_i = ax_i + b$$

DF = n – 2

$$var(a) = \frac{var(\epsilon)}{n\ var(X)}$$

$$var(b) = \frac{var(\epsilon)}{n}\left(1 + \frac{\bar{x}^2}{var(X)}\right)$$

$$var(\hat{y}_i) = \frac{var(\epsilon)}{n}\left(1 + \frac{(x_i-\bar{x})^2}{var(X)}\right)$$

 Online Review Course of Undergraduate Probability and Statistics 18