Finals Cheatsheet

ST2334 Finals Helpsheet

m. zaidan
(for sem1 ay24/25)

Multiplication Principle
If r different experiments are to be performed sequentially with n outcomes, there are $n_1 n_2 \ldots n_r$ number of possible outcomes.

Addition Principles
If an experiment can be done in k different procedures that **do not overlap**, there are $k_1 + k_2 \ldots.$

Permutation
The selection and arrangement of r objects out of n. Order **is taken into consideration**
$P_r^n = \frac{n!}{(n-r)!} = n(n-1)(n-2)..(n-(r-1))$

Combination
The selection of r objects out of n where order is **not taken into consideration**
$\binom{n}{r} = \frac{P_r^n}{P_r^r} = \frac{n!}{r!(n-r)!}$

Inclusion Exclusion Principle

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Conditional probability of B given A: $P(B|A) = \frac{P(A \cap B)}{P(A)}$
Independence: $A \perp B \Leftrightarrow (P(A \cap B) = P(A)P(B)))$
Law of Total Probability: $P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(A_i)P(B|A_i)$
Bayes Theorem: $P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^n P(A_i)P(B|A_k)}$

Discrete Random Variables
Properties:

1. $f(x_i) \geq 0$ for all $x_i \in R_X$
2. $f(x) = 0$ for all $x \notin R_X$
3. $\sum_{i=1}^\infty f(x_i) = 1$ or $\sum_{x_i \in R_X} f(x_i) = 1$

Continuous Random Variables
Properties:

1. **Non-negativity** $f(x) \geq 0$ for all $x \in R_X$, $f(x) = 0$ for $x \notin R_X$
2. **Sum of all probabilities add up to 1** $\int_{R_X} f(x)dx = 1$.
   This particular condition can be represented as $\int_{-\infty}^\infty f(x)dx = 1$
3. For any a, b where $a \leq b$, $P(a \leq X \leq b) = \int_a^b f(x)dx$

Cumulative Distribution Function (cdf)
Properties:

1. Non-decreasing: $x_1 < x_2 \implies F(x_1) \leq F(x_2)$
2. Right-continuous: $F(a) = \lim_{x \to a^+} F(x)$
3. Convergence to 0 and 1 in limits: $\lim_{x \to -\infty} F(x) = 0, \lim_{x \to \infty} F(x) = 1$

Discrete random variables: $\sum_{t \in R_X; t \leq X} P(X = t)$
Continuous random variables: $F(x) = \int_\infty^x f(t)dt$

Expectation

$$\mu_X = E(X) = \int_{-\infty}^\infty x f(x)dx = \int_{x \in R^X} x f(x)dx$$

Properties

1. $E(aX + b) = aE(X) + b$
2. $E(X + Y) = E(X) + E(Y)$
3. $E[g(X)] = \sum_{x \in R_X} g(x)f(x)$ if X is discrete, $E[g(X)] = \int_{R_X} g(x)f(x)$ if X continuous

---

Variance

$$\sigma_x^2 = V(X) = E(X - \mu_X)^2$$
$$V(X) = E(X^2) - [E(X)]^2$$

- $E(X^2) = V(X) + E(X)^2$

With $f(x)$,

- $V(X) = \int_{-\infty}^\infty (x - \mu_X)^2 f(x)dx$ (continuous)
- $V(X) = \sum_{x \in R_X} (x - \mu_X)^2 f(x)$ (discrete)

Distributions

Negative Binomial: Used to find probability of the k^{th} success after n attempts

- Geometric: 1^{st} attempt

Exponential, Geometric: Memoryless $P(X > x | X > y) = P(X > x - y)$

Approximation

Poisson Approximation to Binomial

$X \sim Bin(n, p)$. Suppose that $n \to \infty, p \to 0$ such that $\lambda = np$ remains a constant. Approximately, $X \sim Poisson(np)$.

Good approximation:

$n \geq 20, p \leq 0.05$ or $n \geq 100, np \leq 10$

$$\lim_{p \to 0; n \to \infty} P(X = x) = \frac{e^{-np}(np)^x}{x!}$$

Rule of thumb

$$np > 5, n(1-p) > 5$$

Continuity correction (Binomial to normal)

The continuity correction factor accounts for the fact that a normal distribution is continuous, and a binomial is not.
Generally, it just subtracts or adds 0.5 to the x value.

$$P(x = k) \approx P\left(k - \frac{1}{2} < X < k + \frac{1}{2}\right)$$
$$P(a \leq X \leq b) \approx P\left(a - \frac{1}{2} < X < b + \frac{1}{2}\right)$$
$$P(a < X \leq b) \approx P\left(a + \frac{1}{2} < X < b + \frac{1}{2}\right)$$
$$P(a \leq X < b) \approx P\left(a - \frac{1}{2} < X < b - \frac{1}{2}\right)$$
$$P(a < X < b) \approx P\left(a + \frac{1}{2} < X < b - \frac{1}{2}\right)$$
$$\text{generally,}$$
$$P(x \leq c) \approx P(0 \leq X \leq c) \approx P\left(\frac{-1}{2} < X < c + \frac{1}{2}\right)$$
$$P(x > c) \approx P(c \leq X \leq n) \approx P\left(c + \frac{1}{2} < X < n + \frac{1}{2}\right)$$

Sampling Distributions

Theorem 6

Related to the center and spread of sampling distribution.

For random samples of size n taken from an infinite population with mean μ_X and variance σ_X^2, the sampling distribution of the sample mean $\bar{X}$ has mean μ_X and variance $\frac{\sigma_X^2}{n}$.

$$\mu_{\bar{X}} = E(\bar{X}) = \mu_X \text{ and } \sigma_{\bar{X}}^2 = V(\bar{X}) = \frac{\sigma_X^2}{n}$$

Law of Large Numbers (LLN)

If $X_1, \ldots, X_n$ are independent random variables with the same mean μ and variance σ^2, then for any $\epsilon \in \mathbb{R}$,

$$P(|\bar{X} - \mu| > \epsilon) \to 0 \text{ as } n \to \infty$$

Central Limit Theorem (CLT)

If $\bar{X}$ is the mean of a random sample of size n taken from a population having mean μ and finite variance σ^2, then as $n \to \infty$:

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \to Z \sim N(0, 1)$$

Equivalently, this means:

$$\bar{X} \to N(\mu, \frac{\sigma^2}{n})$$

The CLT states that, under rather general condiitions, for large n, sums and means of random samples drawn from a population follows the normal distribution closely.
(If the random sample comes from a normal population, $\bar{X}$ is normally distributed, regardless.)

Rule of thumb

The mean of a large number of independent samples will have an approximately normal distribution.

- If population is symmetric with no outliers, good approximation to normality appears after as few as 15-20 samples.
- If population is moderately skewed, such as exponential or χ^2, then it can take between 30-50 samples before getting a good approximation
- If population is extremely skewed, CLT may not be appropriate even with a lot of samples.

Distributions

χ^2 Distribution

Let Z be a standard normal random variable. A random variable with the same distribution as Z^2 is called a χ^2 random variable with one degree of freedom.

Let $Z_1, \ldots, Z_n$ be n independent and identically distributed standard normal random variables. A random variable with the same distribution as $Z_1^2 + \ldots + Z_n^2$ is called a x^2 random variable with n degrees of freedom.

We denote a χ^2 random variable with n degrees of freedom as $\chi^2(n)$.

Properties of the χ^2 distribution:

1. If $Y \sim \chi^2(n)$, then $E(Y) = n$, and $V(Y) = 2n$
2. For large n, $\chi^2(n)$ is approximately $N(n, 2n)$
3. If Y_1, Y_2 are independent χ^2 random variables with m, n degrees of freedom respectively, then $Y_1 + Y_2$ is a χ^2 random variable with $m+n$ degrees of freedom.
4. The χ^2 distribution is a family of curves, each determined by degrees of freedom n. All density functions have a long right tail.

Theorem

If S^2 is the variance of a random sample of size n taken from a normal population having the variance σ^2, then the random variable:

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

has a χ^2 distribution with $n-1$ degrees of freedom.

$t-$Distribution

Suppose $Z \sim N(0,1)$ and $U \sim \chi^2(n)$. If Z and U are independent, then

$$T = \frac{Z}{\sqrt{\frac{U}{n}}}$$

follows the t-distribution with n degrees of freedom.

Properties:

- $t-$distribution with n degrees of freedom is denoted $t(n)$
- $t-$ distribution approaches $N(0,1)$ as parameter $\to \infty$. When $n \geq 30$, we can replace it by $N(0,1)$.
- If $T \sim t(n)$, then $E(T) = 0$ and $V(T) = \frac{n}{n-2}$ for $n > 2$.
- Graph of $t-$distribution is symmetric about the vertical axis and resembles the graph of the standard normal distribution.

Theorem 15

If $X_1, \ldots, X_n$ are independent and identically distributed normal random variables with mean μ and variance σ^2 then

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

follows a $t-$distribution with $n-1$ degrees of freedom.

$F-$Distribution

Suppose $U \sim \chi^2(m)$ and $V \sim \chi^2(n)$ are independent. Then the distribution of the random variable

$$F = \frac{\frac{U}{m}}{\frac{V}{n}}$$

is called a $F-$distribution with (m, n) degrees of freedom.

Properties:

- The $F-$distribution with (m, n) degrees of freedom is denoted by $F(m, n)$
- If $X \sim F(m, n)$, then

$$E(X) = \frac{n}{n-2} \text{ for } n > 2$$
$$and$$
$$V(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \text{ for } n > 4$$

- If $F \sim F(n, m)$, then $\frac{1}{F} \sim F(m, n)$. This follows immediately from the definition of the $F-$distribution.
- Values of $F-$distribution can be found in the statistical tables or software. The values of interests are $F(m, n; \alpha)$ such that

$$P(F > F(m, n; \alpha)) = \alpha, F \sim F(m, n)$$

- It can be shown

$$F(m, n; 1-\alpha) = \frac{1}{F(n, m; \alpha)}$$

Estimator

Unbiased estimator

Let $\hat{\Theta}$ be an estimator of θ.
Then, $\hat{\Theta}$ is a random variable based on the sample.

If $E(\hat{\Theta}) = \theta$, $\hat{\Theta}$ is an unbiased estimator of θ.

Maximum error of estimate

$$E = z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$

Confidence Intervals

When $\bar{X} \pm E$ has probability $(1-\alpha)$ of containing μ,

- everytime we take samples and construct the interval estimator, a different confidence interval is computed.
- some confidence intervals contains μ, and some don't.

Since μ is not known,

- there is no way to determine if a confidence interval contains μ or not.
- if the procedure is repeated many times, about $(1-\alpha)$ of the many confidence intervals gotten will contain the true parameter.
  - if we repeat the procedure to get 0.95 confidence intervals, 0.95 of the confidence intervals computed will contain the true parameter.

Errors

Type I error

The rejection of H_0 when H_0 is true.

The probability of a type I error is known as the **significance** of the test.

Significance of the test

$$\alpha = P(\text{type 1 error}) = P(\text{reject } H_0 | H_0 \text{ is true})$$

Type II error

The non-rejection of H_0 when H_0 is false.

The probability of a type II error is referred to as β.

Power of test

$$\beta = P(\text{type 2 error}) = P(\text{do not reject } H_0 | H_0 \text{ is false})$$

Power of the test

The power of the test is the given probability that H_0 is rejected, given that it is false.

Power of test

$$1 - \beta = 1 - P(\text{type 2 error}) = 1 - P(\text{do not reject } H_0 | H_0 \text{ is false})$$

H_0	True	False
Reject	Type I error	CORRECT
Do Not Reject	CORRECT	Type II error

Reducing error

At the same sample size, reducing the type I error results in a higher type II error, and vice-versa.

To reduce both errors, increase the sample size.

Rejection Regions

$\mu \neq \mu_0$: $2P(T < -|t|)$
$\mu < \mu_0$: $P(T < -|t|)$
$\mu > \mu_0$: $P(T > t) = P(T < -|t|)$
