tags: distributionrelationships dist

NOTE:

field: 54

field: $X \sim Gamma(a, b)$ $P(X \leq X) =$

field: $X \sim Gamma(a, b)$ $P(X \leq X) = P(Y \geq a)$ Where $Y \sim Pois(x/b)$

NOTE:

field: 55

field:

$$X_1, \ldots, X_n \sim iid N(0, 1)$$
$$\sum X_i \overset{?}{\sim}$$

field:

$$X_1, \ldots, X_n \sim iid N(0, 1)$$
$$\sum X_i \sim N(0, n)$$

NOTE:

field: 56

field:

$$X_1, \ldots, X_n \sim iid N(\mu_i, \sigma_i^2)$$
$$\sum X_i \overset{?}{\sim}$$

field:

$$X_1, \ldots, X_n \sim iid N(\mu_i, \sigma_i^2)$$
$$\sum X_i \sim N(\sum \mu_i, \sum \sigma_i^2)$$

NOTE:

field: 57

field:

$$X \sim N(\mu, \sigma^2)$$
$$aX + b \stackrel{?}{\sim}$$

field:

$$aX + Y \sim N(a\mu + b, a^2\sigma^2)$$

NOTE:

field: 58

field: $X \sim Binom(1, p) \stackrel{?}{\sim}$

field: $X \sim Bern(p)$

NOTE:

field: 59

field: $X \sim NegBinom(1, p) \stackrel{?}{\sim}$

field: $X \sim Geom(p)$

NOTE:

field: 60

field: $X \sim Gamma(1, \theta) \overset{?}{\sim}$

field: $X \sim Exp(\theta)$

NOTE:

field: 61

field: $X \sim Exp(\theta) \overset{?}{\sim}$

field: $X \sim Gamma(1, \theta)$

NOTE:

field: 62

field: $X \sim Gamma(v/2, 1/2) \overset{?}{\sim}$

field: $X \sim \chi^2(v)$

NOTE:

field: 63

field:

$$X \sim \chi^2(v) \overset{?}{\sim}$$

field:

$$X \sim Gamma(v/2, 2)$$

NOTE:

field: 64

field:

$$X \sim \chi^2(2) \overset{?}{\sim}$$

field:

$$X \sim exp(2)$$

NOTE:

field: 65

field:

$$X \sim Weibull(1, \beta) \overset{?}{\sim}$$

field:

$$X \sim Exp(\beta)$$

NOTE:

field: 66

field: $X_1, X_2 \sim \chi^2(v_i)$ independent $\frac{X_1/v_1}{X_2/v_2}$

field:

$$\frac{(X_1/v_1)}{(X_2/v_2)} \sim F(v_1, v_2)$$

NOTE:

field: 67

field:

$$X \sim beta(1,1) \overset{?}{\sim}$$

field:

$$X \sim Unif(0,1)$$

NOTE:

field: 68

field:

$$X \sim Unif(0,1) \overset{?}{\sim}$$

field:

$$X \sim beta(1,1)$$

NOTE:

field: 69

field: Special case of t

$$X \sim t_1 \overset{?}{\sim}$$

field:

$$X \sim Caucy(0, 1)$$

NOTE:

field: 70

field: Scaled Gamma

$$X \sim Gamma(\alpha, \beta), Y = aX \overset{?}{\sim}$$

field:

$$Y \sim Gamma(\alpha, a\beta)$$

NOTE:

field: 71

field: Scaled Exponential

$$X \sim Exp(\lambda), Y = aX \overset{?}{\sim}$$

field:

$$Y \sim Exp(a\lambda)$$

NOTE:

field: 72

field: Sum of Exponential, equal rate $X_i \sim Exp(\lambda), Y = \sum X_i$

field:

$$Y \sim Gamma(n, \lambda)$$

NOTE:

field: 73

field:

$$X \sim Exp(\lambda), Y = e^{-x}$$

field:

$$Y \sim Beta(\lambda, 1)$$

NOTE:

field: 74

field: Min of Exponential

$$X_1, \ldots, X_n Exp(\lambda_i), Y = \min(X_i) \overset{?}{\sim}$$

field: $Y \sim exp(\sum \lambda_i)$

NOTE:

field: 75

field: Min of Uniform

$$X_i \sim Unif(0,1), Y = \lim n \min(X_i) \overset{?}{\sim}$$

field:

$$Y \sim Exp(1)$$

NOTE:

field: 76

field:

$$X \sim Beta(\alpha, \beta), Y = (1 - X)$$

field:

$$Y \sim Beta(\beta, \alpha)$$

NOTE:

field: 77

field: $X \sim F_X(X), Y = F_X^{-1}(X)$

field: $Y \sim Unif(0,1)$

NOTE:

field: 78

field: $X \sim N(\mu, \sigma^2), Y = e^X$

field: $Y \sim lognormal(\mu, \sigma^2)$

NOTE:

field: 79

field: $X \sim exp(\beta), Y = X^{1/z}$

field: $Y \sim Weibull(z, \beta)$

NOTE:

field: 80

field: Square of Normal $X \sim N(0,1), Y = X^2$

field: $Y \sim \chi^2(1)$

NOTE:

field: 81

field: Square of t $X \sim t(v), Y = X^2$

field: $Y \sim F(1, v)$

NOTE:

field: 82

field: Sum of Poisson $X_i \sim Poisson(\mu_i) Y = \sum X_i$

field: $Y \sim Poisson(\sum \mu_i)$

NOTE:

field: 83

field: Sum of Gamma $X_i \sim Gamma(\alpha_i, \beta), Y = \sum X_i$

field: $Y \sim Gamma(\sum \alpha_i, \beta)$

NOTE:

field: 84

field: Sum of independent Chi-squared $X_i \sim \chi^2(v_i) Y = \sum X_i$

field: $Y \sim \chi^2(\sum v_i)$

NOTE:

field: 85

field: X, Y independent $X, Y \sim N(0,1), X/Y$

field: $X/Y \sim Cauchy(0,1)$

NOTE:

field: 86

field: $X_1, X_2 \sim gamma(\alpha_i, 1)$ independent, $\frac{X_1}{X_1+X_2}$

field:

$$\frac{X_1}{X_1 + X_2} \sim beta(\alpha_1, \alpha_2)$$

NOTE:

field: 87

field: $X_1, X_2 \sim gamma(\alpha_i, \beta_i)$ independent, $\frac{\beta_2 X_1}{\beta_2 X_1 + \beta_1 X_2}$

field:

$$\frac{\beta_2 X_1}{\beta_2 X_1 + \beta_1 X_2} \sim beta(\alpha_1, \alpha_2)$$

NOTE:

field: 88

field: X, Y independent $exp(\mu)$ $X - Y$

field: $X - Y \sim$ double exponential$(0, \mu)$

NOTE:

field: 89

field: $X \sim Gamma(\alpha, \beta)$ $Y = 1/X$

field: Inverted Gamma

NOTE:

field: 90

field: Bernoulii$(p), E(X) =, V(X) =$

field: Bernoulii$(p), E(X) = p, V(X) = p(1-p)$

NOTE:

field: 91

field: Discrete Uniform $N, E(X) =, V(X) =$

field: Discrete Uniform $N, E(X) = \frac{N+1}{2}, V(X) = \frac{(N+1)(N-1)}{12}$

NOTE:

field: 92

field: Cauchy$(\theta, \sigma), E(X) =, V(X) =$

field: Cauchy$(\theta, \sigma), E(X) = na, V(X) = na$

NOTE:

field: 93

field: Double Exponential$(\mu, \sigma), E(X) =, V(X) =$

field: Double Exponential$(\mu, \sigma), E(X) = \mu, V(X) = 2\sigma^2$

NOTE:

field: 94

field: F$(v_1, v_2), E(X) =, V(X) =$

field: F$(v_1, v_2), E(X) = \frac{v_1}{v_2-2}, V(X) = 2\left(\frac{v_2}{v_2-2}\right)^2 \frac{(v_1+v_2-2)}{v_1(v)2-4}$

NOTE:

field: 95

field: Mean and Variance for Distributions not on bible (but in CB)

- Double Exponential$(\mu, \sigma), E(X) =, V(X) =$
- F$(v_1, v_2), E(X) =, V(X) =$
- Logistic$(\mu, \beta), E(X) =, V(X) =$
- Lognormal$(\mu, \sigma^2), E(X) =, V(X) =$
- Pareto$(\alpha, \beta), E(X) =, V(X) =$
- t$(v), E(X) =, V(X) =$
- Weibull$(\gamma, \beta), E(X) =, V(X) =$

field: Mean and Variance. for Distributions not on bible (but in CB)

- Logistic$(\mu, \beta), E(X) = \mu, V(X) = \frac{\phi^2\beta^2}{3}$
- Lognormal$(\mu, \sigma^2), E(X) = e^{\mu+(\sigma^2/2)})V(X) = e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}$
- Pareto$(\alpha, \beta), E(X) = \frac{\beta\alpha}{\beta-1}, V(X) = \frac{\beta\alpha^2}{(\beta-1)^2(\beta-2)}$
- t$(v), E(X) = 0, V(X) = \frac{v}{v-2}$
- Weibull$(\gamma, \beta), E(X) = \beta^{1/\gamma}\Gamma(1+1/\gamma), V(X) = \beta^{2/\gamma}(\Gamma(1+2/\gamma) - \Gamma^2(1+1/\gamma))$

tags: Calculus calc

NOTE:

field: 96

field: $\int_0^\infty e^{-x^2/2} =$

field: $\int_0^\infty e^{-x^2/2} = \sqrt{\pi/2}$

NOTE:

field: 97

field: $\int_0^\infty x^{a-1} e^{-x/b} =$

field: $\int_0^\infty x^{a-1} e^{-x/b} = \Gamma(a) b^a$

NOTE:

field: 98

field: $\int_0^1 x^{a-1}(1-x)^{b-1} =$

field: $\int_0^1 x^{a-1}(1-x)^{b-1} = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

NOTE:

field: 99

field: $\log(x) = y, x =$

field: $\log(x) = y, x = e^y$

NOTE:

field: 100

field: $\lim_{x\to\infty}(1+\frac{a}{x})^x =$

field: $\lim_{x\to\infty}(1+\frac{a}{x})^x = e^a$

NOTE:

field: 101

field: $\lim_{x\to\infty}(1+\frac{a}{x})^x =$

field: $\lim_{x\to\infty}(1+\frac{a}{x})^x = e^a$

NOTE:

field: 102

field: $\frac{d}{dx}f(g(x)) =$

field: $\frac{d}{dx}f(g(x)) = f'(g(x))g'(x)$ (Chain rule)

NOTE:

field: 103

field: $\frac{d}{dx}\int_a^x f(t)dt =$

field: $\frac{d}{dx}\int_a^x f(t)dt = f(x)$ (fundamental theorem of calculus)

NOTE:

field: 104

field: $\int_a^b udv =$
ex: $\int xe^{-x}$

field: $\int_a^b udv = uv|_a^b - \int_a^b vdu$
ex: $u = x, dv = e^{-x}, du = dx, v = -e^{-x}$

$$\begin{aligned}\int xe^{-x} &= -xe^{-x} + \int e^{-x}\\ &= -xe^{-x} - e^{-x} + c\end{aligned}$$

NOTE:

field: 105

field: $\sum_{k=0}^{\infty} \frac{x^k}{k!} =$

field: $\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x$

NOTE:

field: 106

field: $e^x =$

field: $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$

NOTE:

field: 107

field: $\sum_{k=0}^{\infty} x^k =$

field: $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ for $|x| < 1$

NOTE:

field: 108

field: $\sum_{k=0}^{n} x^k =$

field: $\sum_{k=0}^{n} x^k = \frac{1-x^{n+1}}{1-x}$ for $x \neq 1$

NOTE:

field: 109

field: $\lim_{x\to-\infty} e^{-x} =$

field: $\lim_{x\to -\infty} e^{-x} = \infty$

NOTE:

field: 110

field: $\lim_{x\to\infty} e^{-x} =$

field: $\lim_{x\to -\infty} e^{-x} = 0$

NOTE:

field: 111

field:

$$(fg)' =$$

field:

$$(fg)' = f'g + g'f$$

(product rule)

NOTE:

field: 112

field: $\frac{d}{dx}x^n =$

field: $\frac{d}{dx}x^n = nx^{n-1}$

NOTE:

field: 113

field: $\frac{d}{dx}a^x =$

field: $\frac{d}{dx}a^x = a^x ln(a)$

NOTE:

field: 114

field: $\frac{d}{dx}ln(x) =$

field: $\frac{d}{dx}ln(x) = \frac{1}{x}$

NOTE:

field: 115

field: $\frac{d}{dx}(f(x))^n =$

field: $\frac{d}{dx}(f(x))^n = n(f(x))^{n-1}f'(x)$

NOTE:

field: 116

field: $\frac{d}{dx}ln(f(x)) =$

field: $\frac{d}{dx}ln(f(x)) = \frac{f'(x)}{f(x)}$

NOTE:

field: 117

field: $\frac{d}{dx}e^{f(x)} =$

field: $\frac{d}{dx}e^{f(x)} = f'(x)e^{f(x)}$

NOTE:

field: 118

field: $\int x^n =$

field: $\int x^n = \frac{1}{n+1} x^{n+1}$

NOTE:

field: 119

field: $\int \frac{1}{x} =$

field: $\int \frac{1}{x} = ln(|x|)$

NOTE:

field: 120

field: $\int \frac{1}{ax+b} =$

field: $\int \frac{1}{ax+b} = \frac{1}{a} ln(|ax+b|)$

NOTE:

field: 121

field: $\int e^{cx} =$

field: $\int e^{cx} = \frac{1}{c} e^{cx}$

NOTE:

field: 122

field: $\int xe^{-cx^2} =$

field: $\int xe^{-cx^2} = -\frac{1}{2c}e^{-cx^2}$

NOTE:

field: 123

field: U substitution:
example; $\int_1^2 5x^2 \cos(x^3)$

field: $\int_a^b f(g(x))g'(x) = \int_{g(a)}^{g}(b) f(u)du$
Where $u = g(x), du = g'dx$
Ex: $u = x^3, du = 3x^2, x^2du = 1/3du$ $\int_1^2 5x^2 \cos(x^3) = \int_1^8 5/3 \cos(u)du$

NOTE:

field: 124

field: $\Gamma(a) =$

field: $\int_0^\infty t^{a-1}e^{-t}dt$

NOTE:

field: 125

field: $\int_0^\infty t^{a-1}e^{-t}dt$

field: $= \Gamma(a)$

NOTE:

field: 126

field: $\Gamma(a+1) =$

field: $\Gamma(a+1) = a\Gamma(a)$

NOTE:

field: 127

field: $\Gamma(n) =$

field: $\Gamma(n) = (n-1)!$ (for n an integer)

NOTE:

field: 128

field: $\Gamma(1/2) =$

field: $\Gamma(1/2) = \sqrt{\pi}$

NOTE:

field: 129

field: $\Gamma(1) =$

field: $\Gamma(1) = 1$

tags: Theory1

NOTE:

field: 130

field:

	replace	no replacement
number of trials Draw till nth success		

field:

	replace	no replacement
number of trials	Binom	Hypergeometric
Draw till nth success	Nbinom	Negative hypergeometric

NOTE:

field: 131

field: Plug uniform into inverse CDF

field: Get cdf

NOTE:

field: 132

field: Sample Space

field: The set, S, of all possible outcomes of a particular experiment is called the *sample space* for the experiment.

NOTE:

field: 133

field: Event

field: An *event* is any collection of possible outcomes of an experiment, that is, any subset of S (including S itself).

NOTE:

field: 134

field: Union

field: $A \cup B = \{x : x \in A \text{ or } x \in B\}$

NOTE:

field: 135

field: Intersection

field: $A \cap B = \{x : x \in A \text{ and } x \in B\}$

NOTE:

field: 136

field: Complementation

field: $A^c = \{x : x \notin A\}$

NOTE:

field: 137

field: Commutativity

$$A \cup B =$$
$$A \cap B =$$

field: Commutativity

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

NOTE:

field: 138

field: Associativity

$$A \cup (B \cup C) =$$

$$A \cap (B \cap C) =$$

field: Associativity

$$A \cup (B \cup C) = (A \cup B) \cup C$$

$$A \cap (B \cap C) = (A \cap B) \cap C$$

NOTE:

field: 139

field: Distributive Laws

$$A \cap (B \cup C) =$$

$$A \cup (B \cap C) =$$

field: Distributive Laws

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

NOTE:

field: 140

field: DeMorgan's Laws

$$(A \cup B)^c =$$
$$(A \cap B)^c =$$

field: DeMorgan's Laws

$$(A \cup B)^c = A^c \cap B^c$$

$$(A \cap B)^c = A^c \cup B^c$$

NOTE:

field: 141

field: Disjoint

field: Disjoint: Two events A and B are disjoint (or mutually exclusive) if $A \cap B = \emptyset$

NOTE:

field: 142

field:

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) =$$

field:

$$P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\ldots P(A_n|A_1\cdots A_{n-1})$$

NOTE:

field: 143

field:

$$P(A, B, C) =$$

field:

$$P(A, B, C) = P(A)P(B|A)P(C|A, B)$$

NOTE:

field: 144

field:

$$P(A \cup B \cup C) =$$

field:

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

NOTE:

field: 145

field: Pairwise disjoint

field: Two Events A_1, A_2 are pairwise disjoint (or mutually exclusive) if $A_i \cap A_j = \emptyset$ for all $i \neq j$

NOTE:

field: 146

field: Partition

field: If $A_1, A_2, \dots$ are pairwise disjoint and $\cup_{i=1}^{\infty} A_i = S$, then the collection $A_1, A_2, \dots$ forms a partition of S.

NOTE:

field: 147

field: Sigma Algebra

field: A collection of subsets of S is called a sigma algebra (or Borel field), denoted by $\mathcal{B}$, if it satisfies the following three properties:

1. $\emptyset \in \mathcal{B}$ (the empty set is an element of $\mathcal{B}$)
2. If $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$ ($\mathcal{B}$ is closed under complementation)
3. If $A_1, A_2, \dots \in \mathcal{B}$, then $\cup_{i-1}^{\infty} A_i \in \mathcal{B}\mathcal{B}$ is closed under countable unions)

NOTE:

field: 148

field: Probability Function / Kolmogorov Axioms

field: Given a sample space S and an associated sigma algebra $\mathcal{B}$, a probability function is a function P with domain $\mathcal{B}$ that satisfies:

1. $P(A) \geq 0$ for all $A \in \mathcal{B}$
2. $P(S) = 1$
3. If $A_1, A_2, \dots \mathcal{B}$ are pairwise disjoint, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ (Axiom of Countable Additivity)

NOTE:

field: 149

field: If $A \in \mathcal{B}$ and $B \in \mathcal{B}$ are disjoint, then

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

Axiom of Finite Additivity

field: If $A \in \mathcal{B}$ and $B \in \mathcal{B}$ are disjoint, then

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

NOTE:

field: 150

field: Properties of probability functions

1. $P(\emptyset) =$
2. $P(A)$
3. $P(A^c) =$

field: Properties of probability functions

1. $P(\emptyset) = 0$
2. $P(A) \leq 1$
3. $P(A^c) = 1 - P(A)$

NOTE:

field: 151

field: If P is a probability function and A and B are any sets in $\mathcal{B}$, then

$$P(B \cap A^c) =$$

field: If P is a probability function and A and B are any sets in $\mathcal{B}$, then

$$P(B \cap A^c) = P(B) - P(A \cap B)$$

NOTE:

field: 152

field: If P is a probability function and A and B are any sets in $\mathcal{B}$, then

$$P(A \cup B) =$$

field: If P is a probability function and A and B are any sets in $\mathcal{B}$, then

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

NOTE:

field: 153

field: If P is a probability function and A and B are any sets in $\mathcal{B}$, then if $A \subset B$ then

field: If P is a probability function and A and B are any sets in $\mathcal{B}$, then if $A \subset B$ then $P(A) \leq P(B)$

NOTE:

field: 154

field: Bonferroni's Inequality

$$P(A \cap B)$$

field: Bonferroni's Inequality:

$$P(A \cap B) \geq P(A) + P(B) - 1$$

NOTE:

field: 155

field: If P is a probability function, then for any partition $C_1, C_2, \dots P(A) =$

field: If P is a probability function, then for any partition $C_1, C_2, \dots P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$

NOTE:

field: 156

field: Boole's Inequality

$$P(\cup_{i=1}^{\infty} A_i)$$

field: If P is a probability function,

$$P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i) \text{ for any sets } A_1, A_2, \ldots$$

NOTE:

field: 157

field: Fundamental Theorem of Counting

field: Ifa job consists of k separate tasks, the ith of which can be done in n_i ways, $i = 1, \ldots, k$, then the entire job can be done in $n_1 \times n_2 \times \cdots, \times n_k$ ways.

NOTE:

field: 158

field: Ordered without replacement: number of arrangements of size r from n objects

field:

$$\frac{n!}{(n-r)!}$$

eg lottery with $n = 44$ choices for $r = 6$ values, cant use same number twice, order matters

NOTE:

field: 159

field: Unordered without replacement: number of arrangements of size r from n objects

field:

$$\binom{n}{r} = \frac{n!}{r!(n-r!)}$$

eg lottery with $n = 44$ choices for $r = 6$ values, cant use same number twice, order does not matter (Use ordered without replacement and divide by redundant orderings)

NOTE:

field: 160

field: Ordered with replacement: number of arrangements of size r from n objects

field: Ordered with replacement: number of arrangements of size r from n objects

$$n^r$$

eg lottery with $n = 44$ choices for $r = 6$ values, can use same number twice, order matters

NOTE:

field: 161

field: Unordered with replacement: number of arrangements of size r from n objects

field: Unordered with replacement: number of arrangements of size r from n objects

$$\binom{n+r-1}{r} = \frac{(n+r-1)!}{r!(n-1)!}$$

eg lottery with $n = 44$ choices for $r = 6$ values, can use same number twice, order does not matters

NOTE:

field: 162

field: Number of arrangements of size r from n objects

	Without Replacement	With replacement
Ordered		
Unordered		

field: Number of arrangements of size r from n objects

	Without Replacement	With replacement
Ordered	$\frac{n!}{(n-r)!}$	n^r
Unordered	$\binom{n}{r}$	$\binom{n+r-1}{r}$

NOTE:

field: 163

field: Binomial Coefficient $\binom{n}{r}$

field: Binomial Coefficient

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

NOTE:

field: 164

field:

$$P(A|B) =$$

field:

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

NOTE:

field: 165

field: Statistically independent $P(A \cap B) =$

field: Statistically independent $P(A \cap B) = P(A)P(B)$

NOTE:

field: 166

field: If A and B are independent events, what else is independent?

field:

- A and B^c
- A^c and B
- A^c and B^c

NOTE:

field: 167

field: Mutually independent

field: A collection of events $A_1, \dots, A_n$ are mutually independent for any subcollection $A_{i1}, \dots, A_{ik}$, we have

$$P((\cap_{j=1}^{k} A_{ij})) = \prod_{j=1}^{k} P(A_{ij})$$

NOTE:

field: 168

field: Random variable

field: A random variable is a function from a sample space S into the real numbers

NOTE:

field: 169

field: Definition of a pdf

field: A function $f_X(x)$ is a pdf (or pmf) of a random variable X if and only if

1. $f_x(x) \geq 0$ for all x
2. $\sum_x f_x(x) = 1$ or $\int_{-\infty}^{\infty} f_x(x)dx = 1$

NOTE:

field: 170

field: (Theorem) Let X have cdf $F_X(x)$, let $Y = g(X)$

1. If g is an increasing function on X, $F_Y(y) =$ for $y \in Y$
2. If g is a decreasing function on X and X is a continuous random variable, $F_y(y) =$ for $y \in Y$

field: (Theorem) Let X have cdf $F_X(x)$, let $Y = g(X)$

1. If g is an increasing function on X, $F_Y(y) = F_X(g^{-1}(y))$ for $y \in Y$
2. If g is a decreasing function on X and X is a continuous random variable, $F_y(y) = 1 - F_X(g^{-1}(y))$ for $y \in Y$

NOTE:

field: 171

field: Method of pdf

field: Conditions:

1. g is a monotone function
2. $f_X(x)$ is continuous on X
3. $g^{-1}(y)$ has a continuous derivative

Let X have pdf $f_x(x)$ and let $Y = g(Y)$

$$f_Y(y) = f_x(g^{-1}(y))\Big|\frac{d}{dy}g^{-1}(y)$$

NOTE:

field: 172

field: (Theorem) Let X have cdf $F_X(x)$, let $Y = g(X)$

- If g is an increasing function, $F_Y(y) =$
- If g is a decreasing function, and X is a continuous random variable, $F_Y(y) =$

field: (Theorem) Let X have cdf $F_X(x)$, let $Y = g(X)$

- If g is an increasing function, $F_Y(y) = F_x(g^{-1}(y))$
- If g is a decreasing function, and X is a continuous random variable, $F_Y(y) = 1 - F_X(g^{-1}(y))$

NOTE:

field: 173

field: eg: $X \sim Unif(0,1)$, $Y = -log(X)$ $F_Y(y) =$

field: $F_Y(y) = 1 - F_x(g^{-1}(y)) = 1 - F_X(e^{-y}) = 1 - e^{-y}$

NOTE:

field: 173x

field: X is a continuous random variable. For $y > 0$, $Y = X^2$ $F_Y(y) =$

field:

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(X^2 \leq y) \\
&= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\
&= P(X \leq \sqrt{y}) - P(X \leq -\sqrt{y}) \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y})
\end{aligned}$$

NOTE:

field: 174

field: Pdf of $F_X(g(X))$, where $Y = g(X)$

field: Chain rule: $f_Y(y) = g'(y)f(g(y))$

NOTE:

field: 175

field: Method of pdf if g is not monotone all entire domain

field: $f_Y = \sum f_x(g_i^{-1}(y))|\frac{d}{dy}g_i^{-1}(y)|$ $y \in Y$, 0 otherwise
eg: $Y = X^2$,

NOTE:

field: 176

field: $P(Y \leq y)$ when $Y = F_X(x)$

field:

$$\begin{aligned} P(Y \leq y) &= P(X \leq F_x^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) \\ &= y \end{aligned}$$

Y is uniformly distributed

NOTE:

field: 177

field: $M_x(t) =$ (discrete)

field: $M_x(t) = E(e^{tX}) = \sum_x e^{tX} P(X)$ (discrete)

NOTE:

field: 178

field: $M_x(t) =$ (continuous)

field: $M_x(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tX} f_x(x) dx$ (continuous)

NOTE:

field: 179

field: $E(X^n) =$

field: $E(X^n) = M_x^n(0) = \frac{d^n}{dt^n} M_x(t)|_{t=0}$

NOTE:

field: 180

field: $M(aX + b)(t) =$

field: $M(aX + b)(t) = e^{bt} M_x(at)$

NOTE:

field: 181

field: If $E(X^n)$ exists then...

field: If $E(X^n)$ exists then $E(X^m)$ exists for $m \leq n$

NOTE:

field: 182

field: If X_i are independent and $Y = a_1 X_1 + \cdots + a_n X_n + b$, then $M_Y(t) =$

field: If X_i are independent and $Y = a_1 X_1 + \cdots + a_n X_n$, then $M_Y(t) = e^{bt} \prod_{i=1}^{n} M_{X_i}(a_i t)$

NOTE:

field: 183

field: Example of using MGF for finding expected value: MGF gamma: $(\frac{1}{1-\beta t})^{\alpha}$: $E(X) =$

field: $E(X) = \frac{\alpha\beta}{(1-\beta t)^{\alpha+1}}|_{t=0} = \alpha\beta$

NOTE:

field: 184

field: Using MGF to relate distributions: MGF exp $= (1 - \beta t)^{-1}$

field: $Y = \sum X_i$ is gamma as MGF gamma is $(1 - \beta t)^{-\alpha}$

NOTE:

field: 185

field: First step in transforming a RV

field: Determine support

NOTE:

field: 186

field: *nth* Moment of X

field: $E(X^n)$

NOTE:

field: 187

field: nth central moment of X

field: $E[(X - \mu)^n]$

NOTE:

field: 188

field: $(a+b)^n =$

field: $(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}$

NOTE:

field: 189

field: $\sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x} =$

field: $(a+b)^n$

NOTE:

field: 190

field: N balls r red $N-r$ green. Select n balls. Probability that y are red?

field: Hypergeometric distribution(N, r, n)

NOTE:

field: 191

field: Hypergeometric distribution description (N, r, n)

field: N is total balls, r is number red balls, n is number balls selected.

NOTE:

field: 192

field: Negative binomial description

field: Number of Bernoulli trials required to get a fixed number of successes. r being the rth success

NOTE:

field: 193

field: Geometric description

field: Modeling waiting time. X is the trial at which the first success occurs.

NOTE:

field: 194

field: Location-scale family for $f(x)$

field: $\frac{1}{\sigma}f(\frac{(x-\mu)}{\sigma})$

NOTE:

field: 195

field: Given X give the mean and variance for the location-scale random $Y = 1/\sigma f((y-\mu)/\sigma)$ variable

field: $E(Y) = \sigma E(X) + \mu$, $V(Y) = \sigma^2 V(X)$

NOTE:

field: 196

field: $X \sim Pois(\lambda)$ $P(X = x + 1) =$

field: $X \sim Pois(\lambda)$ $P(X = x + 1) = \frac{\lambda}{x+1} P(X = x)$

NOTE:

field: 197

field: $f(y|x) =$ (definition of conditional)

field: $f(y|x) = \frac{f(x,y)}{f_x(x)}$

NOTE:

field: 198

field: $E(g(Y)|x) =$

field: $E(g(Y)|x) = \int_{-\infty}^{\infty} g(y) f(y|x) dy$

NOTE:

field: 199

field: Example of calculating conditional pdfs $f(x, y) = e^{-y}, 0 < x < y < \infty$. $f(y|x) =$

field:

$$\begin{aligned} f_x(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\ &= e^{-x} \end{aligned}$$

$$\begin{aligned} f(y|x) &= \frac{f(x,y)}{f_x(x)} \\ &= \frac{e^{-y}}{e^{-x}} \text{ if } y > x \\ &= \frac{0}{e^{-x}} \text{ if } y \leq x \end{aligned}$$

NOTE:

field: 200

field: Let (X, Y) be given as $f(x, y)$. Then X and Y are independent if

field: Let (X, Y) be given as $f(x, y)$. Then X and Y are independent if there exist functions $g(x), h(y)$ such that $f(x, y) = g(x)h(y)$ (factorization - don't need to compute marginals)

NOTE:

field: 201

field: Let X, Y be independent. Then $E(g(X)h(Y)) =$

field: $E(g(X)h(Y)) = (E(g(X)))(E(h(Y)))$
example: $E(X^2Y) = E(X^2)E(Y)$

NOTE:

field: 202

field: X, Y independent
$Z = X + Y$
$M_Z(t) =$

field: $M_Z(t) = M_x(t)M_Y(t)$

NOTE:

field: 203

field: Method of pdf bivariate

field: $f_{u,v}(u,v) = f_{x,y}(h_1(u,v), h_2(u,v))|J|$
Where $|J| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial v}$
and $u = g_1(x,y), v = g_2(x,y)$ and $x = h_1(x,y), y = h_2(x,y)$

NOTE:

field: 204

field: X, Y independent, $g(X)$ a function only of X and $h(Y)$ a function only of Y. Then

field: $g(X)$ and $g(Y)$ are independent.

NOTE:

field: 205

field: Correlation

field: $\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$

NOTE:

field: 206

field: m independent trials, each trial resulting in one of n outcomes, with probabilities $p_1, \ldots, p_n$. X_i is the count of the number of times the ith outcome occured in the m trials.

field: Multinomial distribution $f(x_1, \ldots, x_n) = \frac{m!}{x_1! \cdots x_n!} p_1^{x_i} \cdots p_n^{x_n}$

NOTE:

field: 207

field: $|E(XY)|$

field: $|E(XY)| \leq E(|XY|) \leq (E(|X|^2))^{1/2}(E(|Y|^2))^{1/2}$ (Cauchy-Schwartz)

NOTE:

field: 208

field: $E(g(X))$ where g is a convex function

field: $E(g(X)) \geq g(E(X))$ where g is a convex function (Jensen's inequlity)

NOTE:

field: 209

field: Ranking of types of means

field: $\mu_{\text{harmonic}} \leq \mu_{\text{geometric}} \leq \mu_{\text{arithmetic}}$ By Jensens inequality (using logs)

NOTE:

field: 210

field: Linear transformations of multivariate normal $X \sim N(\vec{\mu}, \Sigma)$
$A\vec{X} + \vec{b}$

field: $A\vec{X} + \vec{b} \sim N(A\vec{\mu} + \vec{v}, A\Sigma A^t)$

NOTE:

field: 211

field: $X \sim N(\vec{\mu}, \Sigma)$

$\vec{X}_a | \vec{X}_b \sim$

field: $\vec{X}_a | \vec{X}_b \sim N\big(\vec{\mu_a} + \Sigma_{ab}\Sigma_{bb}^{-1}(\vec{x}_b - \vec{\mu}_b), \Sigma_{ba} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$

ex: $(X_1, X_2, X_3), \vec{\mu} = (1, 2, 3)^t, \Sigma = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 3 \end{pmatrix}$ $X1, X_3 | X_2 = 1$
$a = \{1, 3\}, b = \{2\}$
$\mu_a = (1, 3)^t, \mu_b = 1$
$\Sigma_{aa} = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}, \Sigma_{ab} = (1, 1)^t$

NOTE:

field: 212

field: (X, Y) multinomial
$aX + bY \sim$

field: $aX + bY \sim N(a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\rho\sigma_x\sigma_y)$

NOTE:

field: 213

field: (X, Y) multinomial
$Y|X \sim$

field: $Y|X \sim N(\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x), \sigma_Y^2(1 - \rho^2))$

NOTE:

field: 214

field: CDF for Max order statistic

field: $(F(x))^n$

NOTE:

field: 215

field: PDF for Max order statistic

field: $n(F(x))^{n-1}f(x)$

NOTE:

field: 216

field: CDF for Min order statistic

field: $1 - (1 - F(x))^n$

NOTE:

field: 217

field: PDF for Min order statistic

field: $n(1 - F(x))^{n-1} f(x)$

NOTE:

field: 218

field: CDF for kth order statistic

field: $F_{(k)}(x) = \sum_{j=k}^{n} \binom{n}{j} (F(x))^j (1 - F(x))^{n-j}$

NOTE:

field: 219

field: PDF for kth order statistic

field: $f_{(k)}(x) = k\binom{n}{k} f(x) F(x)^{k-1} (1 - F(x))^{n-k}$

NOTE:

field: 1

field: CDF of Geometric (p)

field: $1 - (1 - p)^x$

NOTE:

field: 2

field: CDF of Exponential(β)

field: $1 - e^{-\frac{x}{\beta}}$

NOTE:

field: 3

field:

- $B = \Omega \cap B = (A \cup A^c) \cap B =$
- $P(A^c) =$
- $P(B) =$
- $P(\Omega) =$ $\qquad P(\varnothing) =$
- $(\bigcup_n A_n) =$ $\quad (\bigcap_n A_n) =$ $\qquad$ DeMorgan

field:

- $B = \Omega \cap B = (A \cup A^c) \cap B = (A \cap B) \cup (A^c \cap B)$
- $P(A^c) = 1 - P(A)$
- $P(B) = P(A \cap B) + P(A^c \cap B)$
- $P(\Omega) = 1 \qquad P(\varnothing) = 0$
- $(\bigcup_n A_n) = \bigcap_n A_n \quad (\bigcap_n A_n) = \bigcup_n A_n$ $\qquad$ DeMorgan

NOTE:

field: 4

field: Probability Set intersection

- $P(\bigcup_n A_n) = 1 - P(\bigcap_n A_n^c)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B) \implies P(A \cup B) \leq P(A) + P(B)$
- $P(A \cup B) =$
- $P(A \cap B^c) =$

field: Probability Set intersection

- $P(\bigcup_n A_n) = 1 - P(\bigcap_n \mathsf{A_n})$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  $\implies P(A \cup B) \leq P(A) + P(B)$
- $P(A \cup B) = P(A \cap B^c) + P(A^c \cap B) + P(A \cap B)$
- $P(A \cap B^c) = P(A) - P(A \cap B)$

NOTE:

field: 5

field: $P(A \cap B) =$ when A and Bindependent

field: $P(A \cap B) = P(A)P(B)$ when A and Bindependent

NOTE:

field: 6

field:

$$P(A|B) =$$

field:

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

NOTE:

field: 7

field: Law of total probability

field: Law of total probability

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i) \quad \Omega = \cap_{i=1}^{n} A_i$$

$$P(B) = P(A \cap B) + P(A^c \cap B)$$

NOTE:

field: 8

field: Bayes Theorem $P(A_i|B) =$

field: Bayes Theorem

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)} \quad \Omega = \cup_{i=1}^{n} A_i$$

NOTE:

field: 9

field: CDF Laws

field: CDF Laws

1. Nondecreasing: $x_1 < x_2 \implies F(x_1) \leq F(x_2)$
2. Limits: $\lim_{x \to -\infty} = 0$ and $\lim_{x \to \infty} = 1$
3. Right-Continuous $\lim_{y \to x^+} F(y) = F(x)$

NOTE:

field: 10

field:

$$f_{y|x}(y|x) =$$

field:

$$f_{y|x}(y|x) = \frac{f(x,y)}{f_x(x)}$$

NOTE:

field: 11

field: X, Y independent

- $P(X \leq x, Y \leq y) =$
- $f_{x,y}(x,y) =$

field: X, Y independent

- $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$
- $f_{x,y}(x,y) = f_x(x)f_y(y)$

NOTE:

field: 12

field: Transformations $Z = \phi(X)$

- Discrete: $f_Z(z) =$
- Continuous: $F_Z(z) =$
- Cont, ϕ strictly monotone: $f_z(z)$

field: Transformations $Z = \phi(X)$

- Discrete:

$$f_Z(z) = P(\phi(X) = z) = P(X \in \phi^{-1}(z)) = \sum_{x \in \phi^{-1}(z)} f_x(x)$$

- Continuous (Method of CDF):

$$F_Z(z) = P(\phi(X) \leq z) = \int_{x:\phi(x) \leq z} f(x)dx$$

- Cont, ϕ strictly monotone: (Method of PDF) $f_z(z) = f_x(\phi^{-1}(z))|\frac{d}{dz}\phi^{-1}(z)|$

NOTE:

field: 13

field: Rule of the Lazy Statistician: $E[g(x)] =$

field: Rule of the Lazy Statistician: $E[g(x)] = \int g(x)f_x(x)dx$

NOTE:

field: 14

field: Expectation rules

- $E(c) =$
- $E(cX) =$
- $E(X + Y) =$
- $E(\phi(X)) =$

field: Expectation rules

- $E(c) = c$
- $E(cX) = cE(X)$
- $E(X + Y) = E(X) + E(Y)$
- $E(\phi(X)) \neq \phi(E(X))$

NOTE:

field: 15

field: Conditional expectation

- $E(Y|X = x) =$
- $E(X) =$
- $E(Y + Z|X) =$
- $E(Y|X) = c \implies$

field: Conditional expectation

- $E(Y|X = x) = \int yf(y|x)dy$
- $E(X) = E(E(X|Y))$
- $E(Y + Z|X) = E(Y|X) + E(Z|X)$
- $E(Y|X) = c \implies Cov(X, Y) = 0$

NOTE:

field: 16

field: Variance

- $V(X) = \sigma_x^2 =$
- $V(X + Y) =$
- $V\left[\sum_{i=1}^{n} X_i\right] =$

field: Variance

- $V(X) = \sigma_x^2 = E[(X - E(X))^2] = E(X^2) - E(X)^2$
- $V(X + Y) = V(X) + V(Y) + Cov(X, Y)$
- $V\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} V(X_i) + \sum_{i \neq j} Cov(X_i, X_j)$

NOTE:

field: 17

field: Covariance

- $Cov(X, Y) =$
- $Cov(X, c) =$
- $Cov(Y, X) =$
- $Cov(aX, bY) =$
- $Cov(X + a, Y + b) =$
- $Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) =$

field: Covariance

- $Cov(X, Y) = E[(X - E(X)(Y - E(Y)))] = E(XY) - E(X)E(Y)$
- $Cov(X, c) = 0$
- $Cov(Y, X) = Cov(X, Y)$
- $Cov(aX, bY) = abCov(X, Y)$
- $Cov(X + a, Y + b) = Cov(X, Y)$
- $Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} Cov(X_i, Y_j)$

NOTE:

field: 18

field: Correlation: $\rho(X, Y)$

field: Correlation: $\rho(X, Y) = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$

NOTE:

field: 19

field: Conditional Variance

- $V(Y|X) =$
- $V(Y) =$

field: Conditional Variance

- $V(Y|X) = E\big[(Y - E(Y|X))^2|X\big] = E(Y^2|X) - E(Y|X)^2$
- $V(Y) = E(V(Y|X)) + V(E(Y|X))$

NOTE:

field: 20

field: Law of total probability $k = 2$ (using conditional probability)

field: $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$

NOTE:

field: 21

field: Bayes formula in terms of law of total probability,

field: $P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|B^c)P(B^c)}$

NOTE:

field: 22

field: $P(A \text{ and } B) = P(A, B) =$

field: $P(A \text{ and } B) = P(A|B)P(B) = P(B|A)P(A)$

NOTE:

field: 23

field: Events A and B are independent if

field: $P(A|B) = P(A)$ equivalently $P(A \text{ and } B) = P(A)P(B)$

NOTE:

field: 24

field: Poisson setting

field: The Poisson setting arises in the context of discrete counts of events that occur over space or time with the small probability and where successive events are independent

Eg: 2 on average calls a minute, X is number of calls a minute, $X \sim Pois$

NOTE:

field: 25

field: Poisson approximation of binomial distribution

field: Suppose $X \sim Binom(n, p)$, $Y \sim Pois(\lambda)$. If $n \to \infty$, and $p \to 0$, in such a way that $np \to \lambda > 0$, then for all k, $P(X = k) \to P(Y = k)$. The Poisson distribution with parameter $\lambda = np$ serves as a good approiximation for the binomial distribution when n is large and p is small.

NOTE:

field: 26

field: $E(f(X, Y))$ when X, Y are discrete

field: $E(f(X, Y)) = \sum_x \sum_y f(x, y) P(X = x, Y = y)$

NOTE:

field: 27

field: If X, Y are independent, then $f(X), g(Y)$

field: are also independent

NOTE:

field: 28

field: If X, Y independent, $E(XY) =, E(f(X)g(Y)) =$

field: If X, Y independent, $E(XY) = E(X)E(Y), E(f(X)g(Y)) = E(f(X))E(g(Y))$

NOTE:

field: 29

field: Sum of independent discrete random variables X, Y: $P(X+Y=k)$

field: $P(X+Y=k) = \sum_i P(X=i)P(Y=k-i)$

NOTE:

field: 30

field: $V(X) = 0$

field: If and only if X is a constant

NOTE:

field: 31

field: $E(I_A) =, V(I_A)$ Where I_A is an indicator function

field: $E(I_A) = P(A), V(I_A) = P(A)P(A^c)$

NOTE:

field: 32

field: For discrete jointly distributed random variables,

$$P(X = y|X = x) =$$

field: For discrete jointly distributed random variables,

$$P(X = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

NOTE:

field: 33

field: For discrete random variables $E(Y|X = x) =$

field: For discrete random variables $E(Y|X = x) = \sum_y yP(Y = y|X = x)$

NOTE:

field: 34

field: Problem solving strategy for expected value of counting

field: Use indicator functions for each trial , where $X = \sum I$ and use linearity of expectation

NOTE:

field: 35

field: $P(X > s + t|X > t)$ for geometric, exponential

field: $P(X > s + t|X > t) = P(X > s)$

NOTE:

field: 36

field: Distribution for: A bag of N balls which conatins r red balls and $N-r$ blue balls, X is number of red balls in a sample of size n taken without replacement.

field: Hypergeometric.

NOTE:

field: 37

field: Distribution for modeling arrival time

field: Exponential

NOTE:

field: 38

field: $E(g(X,Y)) =$ (continuous)

field: $E(g(X,Y)) = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} g(x,y)f(x,y)dxdy$

NOTE:

field: 39

field: $Cov(X,Y) =$ (integration)

field: $Cov(X,Y) = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} (x-E(X))(y-E(Y))dxdy$

NOTE:

field: 40

field: Problem solving strategies for functions of random variables

field:

- Methods of cdf: $Y = g(X)$, find cdf $P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y))$
- For finding $P(X < Y)$, set up integrals that cover
- For finding probabilities of independent uniform random variables, use geometric (area) properties

NOTE:

field: 41

field: Quantile

field: If X is a continuous random variable, then the pth quantile is is the number q that satisfies $P(X \leq q) = p/100$

NOTE:

field: 42

field: Poisson process

field: Times between arrivals are modeled as iid exponential random variables with parameter $\lambda = 1/\beta$. Let N_t be the number of arrivals up to time t. Then $N_t \sim Pois(\lambda t)$

NOTE:

field: 43

field: Conditional density function $f_{Y|X}(y|x) =$

field: $f_{Y|X}(y|x) = \frac{f(x,y)}{f_x(x)}$

NOTE:

field: 44

field: Continuous bayes formula

field: $f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_x(x)}{\int_{t=-\infty}^{\infty} f_{Y|X}(y|t)f_x(t)dt}$

NOTE:

field: 45

field: Conditional expectation for continuous random variables $E(Y|X = x)$

field: $E(Y|X = x) = \int_y y f_{Y|X}(y|x)dy$

NOTE:

field: 46

field: Law of total expectation

field: $E(Y) = E(E(Y|X))$

NOTE:

field: 47

field: Properties of conditional expectation

- $E(aY + bZ|X) =$
- $E(g(Y)|X = x) =$
- If X, Y independent, $E(Y|X) =$
- If $Y = g(X)$, then $E(Y|X) =$

field: Properties of conditional expectation

- $E(aY + bZ|X) = aE(Y|X) + bE(Z|X)$
- $E(g(Y)|X = x) = \int_y g(y) f_{Y|X}(y|x) dy$
- If X, Y independent, $E(Y|X) = E(Y)$
- If $Y = g(X)$, then $E(Y|X) = Y$

NOTE:

field: 49

field: Conditional variance $V(Y|X = x)$

field:

$$V(Y|X = x) = \sum_y (y - E(Y|X = x))^2 P(Y = y|X = x)$$

discrete

$$V(Y|X = x) = \int_y (y - E(Y|X = x))^2 f_{Y|X}(y|x) dy$$

continuous

NOTE:

field: 50

field: Properties of conditional variance

- $V(Y|X=x) =$
- $V(aY+b|X=x) =$
- If Y, Z independent, $V(Y+Z|X=x) =$

field: Properties of conditional variance

- $V(Y|X=x) = E(Y^2|X=x) - (E(Y|X=x))^2$
- $V(aY+b|X=x) = a^2V(Y|X=x)$
- If Y, Z independent, $V(Y+Z|X=x) = V(Y|X=x) + V(Z|X=x)$

NOTE:

field: 51

field: $P(X \geq \epsilon)$

field: $P(X \geq \epsilon) \leq E(X)/\epsilon$ (Markov's Inequality)

NOTE:

field: 52

field: $P(|X-\mu| \geq \epsilon)$

field: $P(|X-\mu| \geq \epsilon) \leq \sigma^2/\epsilon^2$ (Chebyshev's inequality, if mean and variance finite)

NOTE:

field: 53

field: $P(\lim_{n\to\infty} S_n/n = \mu) =$

field: $P(\lim_{n\to\infty} S_n/n = \mu) = 1$ (Strong law of large numbers)

tags: TheoryTwo t2

NOTE:

field: 220

field: Definition of Convergence

field: A sequence $\{a_n\}_{n>1}$ of real numbers is said to **converge** to a point $a \in \mathbb{R}$ if for any $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $m > N$ we have $|a_m - a| < \epsilon$

NOTE:

field: 221

field: Example of convergence: $a_n = \frac{1}{n}$

field: For any $\epsilon > 0$, choose N such that $\frac{1}{N} < \epsilon$. Then for any $m > N$ we have that

$$a_n = \frac{1}{n} < \frac{1}{N} < \epsilon$$

and therefore $|a_m - 0| = \frac{1}{n} < \epsilon$

NOTE:

field: 222

field: Given two convergent sequences $\{a_n\}$ and $\{b_m\}$ such that $a_m \to a$ and $b_m \to b$
$\lim_{n\to\infty} a_n b_n =$

field: Given two convergent sequences $\{a_n\}$ and $\{b_m\}$ such that $a_m \to a$ and $b_m \to b$
$\lim_{n\to\infty} a_n b_n = (\lim_{n\to\infty} a_n)(\lim_{n\to\infty} b_n) = ab$

NOTE:

field: 223

field: Definition: Convergence in probability

field: A sequence of random variables $\{X_n\}_{n\geq 1}$ **converges in probability** to a random variable X, if for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$$

We write $X_n \xrightarrow{p} X$

Equivalently, $x_m \xrightarrow{p} x$ if $\lim_{n\to\infty} P(|x_n - x| < \epsilon) = 1$

NOTE:

field: 224

field: Convergence in probability example: Let $\{x_n\}$ be a sequence of random variables such that $x_n \sim N(0, 1/m^2)$
Show that $x_n \xrightarrow{p} 0$:

field: Let $\epsilon > 0$. We obtain $P(|x_n - 0|) = P(x_n > \epsilon) + P(X_n < -\epsilon)$. ie we are looking at the tail probabilities.

Now,

$$\begin{aligned} P(X_n < -\epsilon) + P(x_n > \epsilon) &= P(nx_n < n\epsilon) + P(nx_n > n\epsilon) \\ &= \Phi(n\epsilon) + 1 - \Phi(n\epsilon) \\ &= 2\Phi(-n\epsilon) \underset{n\to\infty}{\to} 0 \end{aligned}$$

Therefore $x_n \xrightarrow{p} 0$

NOTE:

field: 225

field: Example convergence in probability Let $W \sim N(0,1)$ and $U \sim Unif(0,1)$, and define the sequence $\{x_n\}_{n\geq 1}$ as $x_n = W$with prob $1-1/n, U$ with prob $1/n$

Show that $x_n \xrightarrow{p} W$

field: Let $\epsilon > 0$ Then.

$$\begin{aligned} P(|X_n - W| > \epsilon) &= P(|X_n - W| > \epsilon | X_n = W)P(X_n = W) \\ &+ P(|X_n - W| > \epsilon | X_n = U)P(X_n = U) \\ &= 0 \cdot (1 - 1/n) + p_n(1/n) \end{aligned}$$

Where p_n is a probability, and therefore $0 \leq p_n \leq 1$

It follows that $p_n \frac{1}{n} \underset{n\to\infty}{\to} 0$, and therefore $P(|X_n - W| > \epsilon) \underset{n\to\infty}{\to} 0$, for all $\epsilon > 0$, so that $X_n \xrightarrow{p} W$.

NOTE:

field: 226

field: Does $X_n \xrightarrow{p} c$ imply $E(X_n) \to c$?

field: Let $X_n = 0$ with probability $1 - 1/n$, n^2 with probability $1/n$ Then $P(|X_n - 0| > \epsilon) \leq P(X_n = n^2) = 1/n \underset{n\to\infty}{\to} 0$ On the other hand, $E(X_n) = 0 \cdot P(X_n = 0) + n^2 P(X_n = n^2) = 0 + n^2 \frac{1}{n} = n \underset{n\to\infty}{\to} \infty$. Therefore $X_n \xrightarrow{p} c$ does not imply $E(X_n) \to c$

NOTE:

field: 227

field: Does $E(X_n) \to c$ imply $X_n \xrightarrow{p} c$?

field: No.

Let $X_n = 0$, with prob $1 - 1/n$, n with prob $1/n$. Then $E(X_n) = 0 \cdot P(X_n = 0) + nP(X_n = n) = 0 + n1/n = 1$ for all n. But $P(|X_n - 0| > \epsilon) \leq P(X_n = n) = \frac{1}{n} \underset{n\to\infty}{\to} 0$ It follows, $X_n \xrightarrow{p} 0$, and therefore we have $E(X_n) \to c$does not imply $X_n \xrightarrow{p} c$

NOTE:

field: 228

field: Suppose $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ be two sequences of random variables such that $X_n \xrightarrow{p} x_0$ and $Y_n \xrightarrow{p} y_0$ as $n \to \infty$, where $x_o, y_0 \in \mathbb{R}$

What properties do we have?

field:

- $X_n \pm Y_m \xrightarrow{p} x_0 \pm y_0$ as n increases to ∞
- $X_n Y_n \xrightarrow{p} x_0 y_0$ as n increases to ∞
- $X_n / Y_n \xrightarrow{p} x_0 / y_0$ as n increases to infinity, provided that $P(Y_n = 0) = 0$ fro all n and $y_0 \neq 0$

NOTE:

field: 229

field: Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables such that $x_n \xrightarrow{p} x_0 \in \mathbb{R}$, as $n \to \infty$, and let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function . Then

$$g(X_n) \xrightarrow{p} \text{ as } n \to \infty$$

field: Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables such that $x_n \xrightarrow{p} x_0 \in \mathbb{R}$, as $n \to \infty$, and let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function . Then

$$g(X_n) \xrightarrow{p} g(x_0) \text{ as } n \to \infty$$

NOTE:

field: 230

field: Proof of: Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables such that $x_n \xrightarrow{p} x_0 \in \mathbb{R}$, as $n \to \infty$, and let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function . Then

$$g(X_n) \xrightarrow{p} g(x_0) \text{ as } n \to \infty$$

field: Since g is continuous at $X = x_0$, we have that for any $\epsilon > 0$, there exits $\delta > 0$ such that $|g(x) - g(x_0)| > \epsilon$ implies $|x - x_0| > \delta$

We obtain

$$0 \leq P(|g(X_n) - g(x_0)| > \epsilon) \leq P(|X_n - x_0| > \delta) \underset{n\to\infty}{\to} 0$$

NOTE:

field: 231

field: Weak Law of Large numbers

field: Let $X_1, X_2, X_3 \ldots$ Be a sequence of iid random variables with $E(X_1) = \mu$ (finite) and $V(X_1) = \sigma^2 < \infty$, and define $\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (the sample mean).

Then

$$\bar{X_n} \xrightarrow{p} \mu \text{ as } n \to \infty$$

NOTE:

field: 232

field: Proof of Weak Law of Large Numbers

field:

$$\begin{aligned} P(|\bar{X}_n - \mu| > \epsilon) &= P((\bar{X}_n - \mu)^2 > \epsilon^2) \\ &\leq \frac{E((\bar{X}_n - \mu)^2)}{\epsilon^2} \text{ by Chebyshev's Inequality} \\ &= \frac{V(\bar{X}_n)}{\epsilon^2} \text{ by def of variance} \\ &= \frac{\sigma^2}{n\epsilon^2} \underset{n\to\infty}{\to} 0 \end{aligned}$$

Therefore $\bar{X}_n \xrightarrow{p} \mu$

NOTE:

field: 233

field: Consistency

field: If our estimate converges in probability to the value of the parameter of interest as the sample size n increases

NOTE:

field: 234

field: Consistency of S^2

field: Suppose $X_1, X_2, \ldots$ is a sequence of iid random varialbes with $E(X_1) = \mu$ finite and $V(X_1) = \sigma^2 < \infty$

and define

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \quad \text{The sample variance}$$

Can we show that S_n^2 is a consistent estimate of σ^2? In other words, can we show taht $S_n^2 \xrightarrow{p} \sigma^2$ as $n \to \infty$

Using Chebychev's inequality, we obtain

$$\begin{aligned} P(|S_n^2 - \sigma^2| > \epsilon) &\leq \frac{E[(S_n^2 - \sigma^2)^2]}{\epsilon^2} \\ &= \frac{V(S_n^2)}{\epsilon^2} \end{aligned}$$

There fore, a sufficient condition that S_n^2 converges in probablility to σ^2 is that the variance of S_n^2 $V(S_n^2) \to 0$, as $n \to \infty$

NOTE:

field: 235

field: $V(S_n^2) \to 0$ as long as

field: $V(S_n^2) \to 0$ as long as the fourth central moment $\mu_4 = E[(X_1 - \mu)^4]$ is finite.

NOTE:

field: 236

field: Khinchin's WLLN

field: Let $X_1, X_2, \ldots$ be a sequence of iid random variables with $E(X_1) = \mu$ (finite). Then, $\bar{X}_n \xrightarrow{p} \mu$ as $n \to \infty$

NOTE:

field: 237

field: Let $X_1, X_2 \ldots$ be a sequence of random variables, such that for some $r > 0$ and $c \in \mathbb{R}$, $E[|X_n - c|^r] \underset{n\to\infty}{\to} 0$. Then $X_n \xrightarrow{p}$, as $n \to \infty$

field: (A general result to establish convergence in probability)

Let $X_1, X_2 \ldots$ be a sequence of random variables, such that for some $r > 0$ and $c \in \mathbb{R}$, $E[|X_n - c|^r] \underset{n\to\infty}{\to} 0$. Then $X_n \xrightarrow{p} c$, as $n \to \infty$

NOTE:

field: 238

field: Consistent estimator for $X_1, X_2, \ldots X_n \sim$ iid Univorm$(0, \theta)$, $\theta > 0$. (and sketch of proof)

field: $X_{(n)} = \max(X_1, \ldots X_n)$ (the largest order statistic)

Proof

First recall that the pdf of $X_{(n)}$ is given by

$$f(x) = nx^{n-1}\theta^{-n}, 0 < x < \theta, 0\text{otherwise}$$

We obtain

$$
\begin{aligned}
E(X_{(n)}) &= \int_0^\theta x f(x)dx \\
&= n\theta^{-n}\int_0^\theta x^n dx \\
&= \frac{n}{n-1}\theta \\
E(X_{(n)}^2) &= \int_0^\theta x^2 f(x)dx \\
&= n\theta^{-n}\int_0^\theta x^{n+1} dx \\
&= \frac{n}{n+2}\theta^2
\end{aligned}
$$

We have

$$
\begin{aligned}
E[(X_{(n)}-\theta)^2] &= E(X_{(n)}^2) - 2\theta E(X_{(n)}) + \theta^2 \\
&= \frac{n}{n+2}\theta^2 - 2\theta\frac{n}{n+1}\theta + \theta^2 \\
&\cdots \\
&= \frac{2\theta^2}{(n+1)(n+2)} \underset{n\to\infty}{\to} 0
\end{aligned}
$$

Hence, taking $c = 0$ and $r = 2$, from the previous theorem, we obtain $X_{(n)} \xrightarrow{p} \theta$ as $n \to \infty$

NOTE:

field: 239

field: Definition Almost Sure Convergence

field: A sequence $\{X_n\}_{n\geq 1}$ of random variables is said to converge **Almost Surely** to a random variable X if for every $\epsilon > 0$,

$$
P(\lim_{n\to\infty} |X_n - X| > \epsilon) = 0
$$

We write $X_n \xrightarrow{a.s} X$ as $n \to \infty$

NOTE:

field: 240

field: Strong Law of Large Numbers

field: Let $X_1, X_2, \ldots$ be an iid sequence of random variables, with $E(X_1) = \mu$ (finite) and $V(X_1) = \sigma^2 < \infty$. Then,

$$\bar{X_n} \stackrel{a.s}{\to} \mu \quad \text{as } \mu \to \infty$$

NOTE:

field: 241

field: Does convergence in probability imply convergence almost surely?

field: No. Let $\Omega = [0.1]$, with uniform probability distribution. Define the sequence $\{X_n\}_{n\geq 1}$ as:

$$\begin{aligned}
X_1(\omega) &= \omega + \mathbb{I}_{[0,1]}(\omega) \\
X_2(\omega) &= \omega + \mathbb{I}_{0,1/2}(\omega) \\
X_3(\omega) &= \omega + \mathbb{I}_{1/2,1}(\omega) \\
X_4(\omega) &= \omega + \mathbb{I}_{0,1/3}(\omega) \\
X_5(\omega) &= \omega + \mathbb{I}_{1/3,2/3}(\omega) \\
&\vdots
\end{aligned}$$

$X_5(\omega) = \omega + 1$

Let $X(\omega) = \omega$, then it is easy to show that $X_n \stackrel{p}{\to} X$ because $P(|X_n - X| \geq \epsilon) = P([a_n, b_n])$, where $l_n = \text{length}([a_n, b_n]) \underset{n\to\infty}{\to} 0$.

However X_n does not converge to X almost surely, because for every $\omega \in [0, 1]$, alternates between ω and $\omega + 1$, infinetly often as $n \to \infty$

NOTE:

field: 242

field: Convergence in Distribution

field: A sequence $\{X_n\}_{n\geq 1}$ of random variables converges in distribution to a random variable X if,

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at all points x where $F_X(x)$ is continuous
We write $X_n \xrightarrow{d} X$

NOTE:

field: 243

field: Example of convergence in distribution
Let $X_n \sim N(0, \frac{n+1}{n})$, and $X \sim N(0,1)$. We want to show that $X_n \xrightarrow{d} X$.

field:

$$\begin{aligned} P(X_n \leq X) &= P(\sqrt{\frac{n}{n+1}} X_n \leq \sqrt{\frac{n}{n+1}} x) \\ &= \Phi(\sqrt{\frac{n}{n+1}} x) \underset{n\to\infty}{\to} \Phi(x) \end{aligned}$$

And we obtain that $F_{X_n} \to \Phi(x) = F_X(x), \forall x$, and therefore $X_n \xrightarrow{d} X$

NOTE:

field: 244

field: Does Convergence in probability imply convergence in distribtuion?

field: Yes

NOTE:

field: 245

field: Does Convergence in distribution imply convergence in probability?

field: No - unless converges in distribution to a constant

NOTE:

field: 246

field: A sequence $\{X_n\}_{n\geq 1}$ of random variables converges in probability to a constant $c \in \mathbb{R}$ if and only if

field: A sequence $\{X_n\}_{n\geq 1}$ of random variables converges in probability to a constant $c \in \mathbb{R}$ if and only if the sequence converges in distribution to c

NOTE:

field: 247

field: If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$ we have that

1. $X_n \pm Y_n$
2. $X_n Y_n$

field: In general it is not true that if $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$ we have that

1. $X_n \pm Y_n \xrightarrow{d} X + Y$
2. $X_n Y_n \xrightarrow{d} XY$

NOTE:

field: 248

field: Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables such that $X_n \xrightarrow{d} X$, for some random variable X (possibly a constant). Then for any continuous funciton $g : \mathbb{R} \to \mathbb{R}$, we have $g(X_n) \xrightarrow{d}$

field: Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables such that $X_n \xrightarrow{d} X$, for some random variable X (possibly a constant). Then for any continuous funciton $g : \mathbb{R} \to \mathbb{R}$, we have $g(X_n) \xrightarrow{d} g(X)$

NOTE:

field: 249

field: Let $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ be two sequences of random variables such that $X_n \xrightarrow{d} X$ for some random variable X (possibly a constant) and $Y_n \xrightarrow{p} c \in \mathbb{R}$

Then, as $n \to \infty$,

1. $X_n \pm Y_n \xrightarrow{d}$
2. $X_n Y_n \xrightarrow{d}$
3. $X_n / Y_n \xrightarrow{d}$ provided $P(Y_n = 0) = 0 \forall n$ and $c \neq 0$

field: Slutsky's Theorem Let $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ be two sequences of random variables such that $X_n \xrightarrow{d} X$ for some random variable X (possibly a constant) and $Y_n \xrightarrow{p} c \in \mathbb{R}$

Then, as $n \to \infty$,

1. $X_n \pm Y_n \xrightarrow{d} X \pm c$
2. $X_n Y_n \xrightarrow{d} cX$
3. $X_n / Y_n \xrightarrow{d} X/c$ provided $P(Y_n = 0) = 0 \forall n$ and $c \neq 0$

NOTE:

field: 250

field: Central Limit Theorem

field: Let $X_1, X_2, \ldots$ be an iid sequence of random variables, with $E(X_1) = \mu$(finite) and $V(X_1) = \mu^2 < \infty$

Then, for $\bar{X_n} = \frac{1}{n}\sum_{i=1}^{\infty} X_i$ (the sample mean), we have that

$$\frac{\sqrt{n}(\bar{X_n} - \mu)}{\sigma} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty$$

NOTE:

field: 251

field: Equivalent results of CLT

field:

- $\frac{(\bar{X_n} - \mu)}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0,1)$
- $\sqrt{n}(\bar{X_n} - \mu) \xrightarrow{d} N(0,\sigma^2)$
- $\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0,1)$
- $\bar{X_n} \xrightarrow{d} N(\mu, \sigma^2/n)$

NOTE:

field: 252

field: Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables such that the mgf $M_{X_n}(t)$ of X_n exists in a neighborhood of 0, for all , and suppose that

$$\lim_{n\to\infty} M_{X_n}(t) = M_X(t) \quad \text{for all } t \text{ in a neighborhood of } 0$$

where $M_X(t)$ is the mgf for some random variable X. Then,

field: Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables such that the mgf $M_{X_n}(t)$ of X_n exists in a neighborhood of 0, for all , and suppose that

$$\lim_{n\to\infty} M_{X_n}(t) = M_X(t) \quad \text{for all } t \text{ in a neighborhood of } 0$$

where $M_X(t)$ is the mgf for some random variable X. Then, there exists a unique cdf $F_x(x)$ whose moments are determined by $M_y(t)$ and for all x, where $F_x(x)$ is continuous we have $\lim_{n\to\infty} F_{X_n}(x) = F_x(x)$

NOTE:

field: 253

field: $\frac{\sqrt{n}(\bar{X}-\mu)}{S_n} \xrightarrow{d}$

field: Using the CLT, and slutsky's theorem, we have

$$\frac{\sqrt{n}(\bar{X}-\mu)}{S_n} = \frac{\sigma}{S_n} \cdot \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \xrightarrow{d} N(0,1)$$

NOTE:

field: 254

field: $g(X) \approx$
$E(g(X)) \approx$, $V(g(X)) \approx$

field:

$$g(X) \approx g(\mu) + g'(X)(X-\mu)$$

Using a first order taylor approximation $E(g(X)) \approx g(\mu)$, $V(g(X)) \approx [g'(\mu)]^2 V(X)$

NOTE:

field: 255

field: Delta Method

field: Let $\{Y_n\}_{n\geq 1}$ be a sequence of random variables such that $\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$ as $n \to \infty$. Suppose that for a given function g and a specific value of θ, $g'(\theta)$ exists and is not equal to zero. Then

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{d} N(0, \sigma^2[g'(\theta)]^2)$$

as $n \to \infty$

NOTE:

field: 256

field: Second Order delta method

field: Let $\{Y_n\}_{n\geq 1}$ be a sequence of random variables such that $\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$ as $n \to \infty$. And that for a given function g as specific value of θ, we hvae $g'(\theta) = 0$, but $g''(\theta)$ Exists and is not equal to 0. Then

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{d} \sigma^2 \frac{g''(\theta)}{2} \chi_1^2 \quad \text{as } n \to \infty$$

NOTE:

field: 257

field: $\chi_n^2 \dot{\sim}$ for sufficiently large n

field: $\chi_n^2 \dot{\sim} N(n, 2n)$

NOTE:

field: 258

field: Definition Statistic

field: Let $X_1, \dots, X_n$ be a random sample from a given population. Then, any observable real-valued (or vector-valued) function $T(\mathbf{X}) = T(X_1, \dots, X_n)$ of the random variables $X_1, \dots, X_n$ is called a **Statistic**

NOTE:

field: 259

field: Sampling Distribution

field: The probability distribution of the statitic $T(\mathbf{X})$ is called the **Sampling Distribution** of $T(\mathbf{X})$

NOTE:

field: 260

field: Sufficient Statistic

field: A statistic $T(\mathbf{X})$ is a **Sufficient Statistic** for θ, if the conditional distribution of the sample $\mathbf{X}$ given the value of $T(\mathbf{X})$ does not depend on θ

NOTE:

field: 261

field: Determine if $T(\mathbf{X}) = \sum X_i$ where $X_i \sim Bern(p)$ is sufficient for p using definition of sufficiency

field:

$$
\begin{aligned}
P(\mathbf{X}=\mathbf{x}\big|T=t) &= \frac{P(\cap_{i=1}^{n} X_i = x_i)}{P(T=t)} \\
&= \prod_{i=1}^{n} \frac{P(X_i = x_i)}{P(T=t)} \quad \text{by independence} \\
&= \frac{p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}}{\binom{n}{t} p^t (1-p)^{n-t}} \quad \text{Because } T \sim \text{Binom}(n,p) \\
&= \frac{p^t(1-p)^{n-t}}{\binom{n}{t} p^t (1-p)^{n-t}} \quad \text{because } t = \sum_{i=1}^{n} x_i \\
&= \frac{1}{\binom{n}{t}} \quad \text{which is free of } p
\end{aligned}
$$

NOTE:

field: 262

field: How to show sufficiency (not using factorization)

field: Let $p(\mathbf{X}|\theta)$ be the joint PDF or PMF of $\mathbf{X}$ and $q(t|\theta)$ the PDF or PMF of the statistic $T(\mathbf{X})$. Then $T(\mathbf{X})$ is a sufficient statistic for θ if for every $\mathbf{X}$ in the sample space, the ratio

$$\frac{p(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)}$$

is constant as a function of θ

NOTE:

field: 263

field: Suppose that $X_1, \ldots X_n$ are iid $N(\mu, \sigma^2)$ where σ^2 is known. If the statistic $T(\mathbf{X}) = \bar{X}_n$ sufficient for μ?

field:

$$\frac{f(\mathbf{x}|\mu)}{q(T(\mathbf{X})|\mu)} = \frac{(2\pi\sigma^2)^{n/2} e^{-\frac{1}{2\sigma^2}[\sum_{i=1}^n (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2]}}{(2\pi\sigma/n)^{-1/2} e^{-\frac{1}{2\sigma^2}(\bar{x}-\mu)^2}}$$
$$= n^{-1/2}(2\pi\sigma^2)^{-(n-1)/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\bar{x})^2}$$

Which does not depend on μ, and therefore $\bar{X}_n$ is suffient for μ as long as σ^2 is known

NOTE:

field: 264

field: The joint pdf of the sample $\mathbf{X} = (X_1, X_2, \ldots X_n)$ isSuppose that $X_1, \ldots X_n$ are iid $N(\mu, \sigma^2)$ where σ^2 is known.

field:

$$f(\mathbf{x}|\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}(x_i-\mu)^2}$$
$$= (2\pi\sigma^2)^{n/2} e^{-1/2\sigma^2 \sum_{i=1}^n (x_i-\mu)^2}$$
$$= (2\pi\sigma^2)^{n/2} e^{-1/2\sigma^2 \sum_{i=1}^n (x_i-\bar{x}+\bar{x}-\mu)^2}$$
$$= (2\pi\sigma^2)^{n/2} e^{-1/2\sigma^2 \sum_{i=1}^n (x_i-\bar{x})^2 + 2(\bar{x}-\mu)\sum_{i=1}^n (x_i-\bar{x}) + n(\bar{x}-\mu)^2}$$
$$= (2\pi\sigma^2)^{n/2} e^{-1/2\sigma^2 (\sum_{i=1}^n (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2)}$$

NOTE:

field: 265

field: Show a statistic $T(\mathbf{X})$ is sufficient

field: Neyman factorization theorem Let $f(\mathbf{x}|\theta)$ denote the joint pdf or pmf of the sample $\mathbf{X}$, A statistic $T(\mathbf{X})$ is a sufficient statistic for θ if and only if there exists functions $g(t|\theta)$ and $h(\mathbf{x})$ such that for all sample points $\mathbf{x}$ and all values of θ we can write

$$f(\mathbf{x}|\theta) = g(T(x)|\theta)h(\mathbf{x})$$

Note, in the theorem

- The function $g(T(\mathbf{X})|\theta)$ depends on $\mathbf{x} = (x_1, \ldots x_n)$ only through the statistic $T(\mathbf{X})$.
- The function $h(\mathbf{X})$ does not depend on θ

NOTE:

field: 266

field: Exponential Family

field:

$$f(\mathbf{X}|\theta) = \mathbf{h}(\mathbf{x})\mathbf{c}(\theta)\mathbf{e}^{\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{w_i}((\theta))\mathbf{t_i}(\mathbf{x})}$$

NOTE:

field: 267

field: Sufficiency in the exponential family

field: Let $X_1, \ldots, X_n$ be iid observations from a PDF or PMF, $f(x|\boldsymbol{\theta})$ that belongs to an exponential family of the form

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})e^{\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)}$$

Where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$, $d \leq k$. Then

$$T(\mathbf{X}) = \big(\sum_{j=1}^{k} t_i(x_j), \cdots, \sum_{j=1}^{k} t_k(x_j)\big)$$

NOTE:

field: 268

field: Minimal Sufficient Statistic

field: A sufficient statistic $T(\mathbf{X})$ is called a **Minimal Sufficient Statistic** if for any other sufficient statistic $T'(\mathbf{X})$, $T(\mathbf{X})$ is a function of $T'(\mathbf{X})$

NOTE:

field: 269

field: Determining if a statistic is minimal sufficient

field: Let $f(x|\theta)$ be the PDF or PMF of a sample $\mathbf{X}$. Suppose there exists a function $T(x)$ such that, for every two sample points, $\mathbf{x}$ and $\mathbf{y}$, the ratio $\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)}$ is constant as a fucntion of θ iff and only if $T(\mathbf{x}) = T(\mathbf{y})$. Then $T(\mathbf{x})$ is a minimal sufficient statistic for θ.

NOTE:

field: 270

field: Example of finding a minimal sufficient statistic: Suppose that $X_1, \ldots, X_n$ are idd Bernoulli(p). What is a minimal sufficient statistic for p?

field:

$$\begin{aligned} f(\mathbf{x}|p) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\ &= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \end{aligned}$$

And therefore for any two sample points $\mathbf{x}$ and $\mathbf{y}$, we obtain

$$\frac{f(\mathbf{x}|p)}{f(\mathbf{y}|p)} = \frac{p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}}{p^{\sum_{i=1}^n y_i}(1-p)^{n-\sum_{i=1}^n y_i}}$$
$$= p^{\sum_{i=1}^n x_i - \sum_{i=1}^n y_i}(1-p)^{\sum_{i=1}^n y_i - \sum_{i=1}^n x_i}$$

Which is constant as a function of p iff $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$

Hence it follows from Lehman-Sheffe that $T(\mathbf{x}) = \sum_{i=1}^n x_i$ is minimal sufficient for p

NOTE:

field: 271

field: Minimal sufficient statistic for μ, σ^2, where the Xs are $N(\mu, \sigma^2)$

field: $T(\mathbf{x}) = (\bar{x}, S_x^2)$ by Lehmann-Schaffe is minimal sufficient.

NOTE:

field: 272

field: Facts about sufficiency

field:

- The entire sample $\mathbf{X}$ is always sufficeint.
- Any one-to-one funciton of a minimal sufficient statisitc is also a minimal sufficient statistic

NOTE:

field: 273

field: Ancillary Statistic

field: A statistic $S(\mathbf{X})$ whose distribution does not depend on the parameter θ is called an ancillary statistic for θ

NOTE:

field: 274

field: Complete statistic

field: Let $f(t|\theta)$ be the family of pdf's or pmfs for a statistic $T = T(\mathbf{x})$.

The family of probability distributions is called **complete** (with respect to θ) if $E_\theta(g(t)) = 0$ for all θ, implies $P_\theta(g(T) = 0) = 1$ for all θ

Equivalently, we say that $T = T(\mathbf{X})$ is a complete statistic.

In short, a statistic $T = T(\mathbf{x})$ is complete, if $E_\theta(g(T)) = 0$ for all θ implies $g(t) = 0$ with probability 1

NOTE:

field: 275

field: (Binomial complete sufficient statistic)

field: Suppose the statistic $T \sim Binom(n, p)$, $0 < p < 1$, and let g be a function such that $E_p(g(T)) = 0$ for all p.

Then, with $r = (\frac{p}{1-p})^t$

$$
\begin{aligned}
0 &= E_p(g(T)) \\
&= \sum_{t=0}^{n} g(t)\binom{n}{t} p^t (1-p)^{n-1} \\
&= (1-p)^n \sum_{t=0}^{n} g(t)\binom{n}{t} (\frac{p}{1-p})^t \\
&= (1-p)^n \sum_{t=0}^{n} g(t)\binom{n}{t} r^t \\
&= \neq 0 \cdot \text{This is a polynomial of degree } n \text{ in } r \text{ with coefficients } g(t)\binom{n}{t}
\end{aligned}
$$

For the polynomial to be 0 for all r (and consequently for all p) each coefficient must be zero and therefore it must be the case that $g(t) = 0$ for $t = 0, 1, 2, \cdots, n$ Since $T \sim Binom(n, p)$, we have that T takes on the values $t = 0, 1, 2, \ldots n$ with probability 1 and therefore, we obtain $P_p(g(T) = 0) = 1$. Hence T is a complete statistic.

NOTE:

field: 276

field: Uniform$(0, \theta)$ complete sufficient statistic

field: $T(\mathbf{X}) = X_{(n)}$

Suppose that $X_1, \ldots, X_n$ are iid Uniform$(0, \theta)$, $\theta > 0$. We know that $T(\mathbf{X}) = X_{(n)}$ (the max order statistic) is sufficient for θ. Furtheremore ,

$$f(t|\theta) = nt^{n-1}\theta^{-n} \quad 0 < t < \theta$$

Now suppose that $g(t)$ is a function satisfying $E_\theta(g(T)) = 0, \forall\theta$ Differentiating on both sides with respect to θ,

$$
\begin{aligned}
0 &= \frac{d}{d\theta}E_\theta(g(t)) \\
&= \frac{d}{d\theta}\int_0^\theta g(t)nt^{n-1}\theta^{-n}dt \\
&= \theta^{-n}\frac{d}{d\theta}\int_0^\theta g(t)nt^{n-1}dt + (\frac{d}{d\theta}\theta^{-n})\int_0^\theta g(t)nt^{n-1}dt \\
&= \theta^{-n}g(\theta)n\theta^{n-1} + 0
\end{aligned}
$$

Since $n\theta^{-1} \neq 0$, we must have that $g(\theta) = 0 \quad \forall\theta > 0$. And therefore T is complete.

NOTE:

field: 277

field: Does minimal sufficent imply complete?

field: No

Suppose that $X_1, \ldots X_n$ are iid $N(\theta, \theta^2)$ where $\theta \in \mathbb{R}$ is the unknown parameter of interest.

We have

$$
\begin{aligned}
\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} &= \frac{(2\phi\sigma^2)^{-n/2}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\theta)^2}}{(2\phi\sigma^2)^{-n/2}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(y_i-\theta)^2}} \\
&= \frac{e^{-\frac{1}{2\sigma^2}[\sum_{i=1}^n x_i^2 - 2\theta\sum_{i=1}^n x_i]}}{e^{-\frac{1}{2\sigma^2}[\sum_{i=1}^n y_i^2 - 2\theta\sum_{i=1}^n y_i]}}
\end{aligned}
$$

Which is free of θ if $\sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2$ and $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$

It follows that $T(\mathbf{X}) = (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ is minimal sufficient for θ

Now observe that $T_1(\mathbf{X}) = \sum_{i=1}^n x_i \sim N(n\theta, n\theta^2)$ and therefore

$$
\begin{aligned}
E(T_1^2) &= V(T_1) + [E(T_1)]^2 \\
&= n\theta^2 + n^2\theta^2 \\
&= n\theta^2(1+n)
\end{aligned}
$$

On the other hand, for $T_2 = \sum_{i=1}^n x_i^2$,

$$\begin{aligned} E(T_2) &= nE(X_1)^2 \\ &= n[V(X_1) + [E(X_1)]^2] \\ &= n\theta^2 + n\theta^2 \\ &= 2n\theta^2 \end{aligned}$$

Then, taking $h(t_1, t_2) = 2t_1^2 - (n+1)t_2$, we have

$$\begin{aligned} E_\theta[h(T_1, T_2)] &= E_\theta[2T_1^2 - (n+1)T_2] \\ &= 2E_\theta(T_1^2) - (n+1)E(T_2) \\ &= 2n(n+1)\theta^2 - 2n(n+1)\theta^2 \\ &= 0 \quad \forall\theta \end{aligned}$$

But because $h(\mathbf{t}) \neq 0 \quad \forall\theta$, we have that $T(\mathbf{X})$ is not complete.

NOTE:

field: 278

field: Complete statistics in the exponential family

field: Let $X_1, \ldots, X_n$ be iid observations from an exponential family. with PDF or PMF of the form

$$f(x|\theta) = h(x)c(\theta)e^{\sum_{j=1}^k \omega_j(\theta_j)t_j(x)}$$

Where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$

Then, the statistic $T(\mathbf{X}) = (\sum_{i=1}^n t_1(x_i), \sum_{i=1}^n t_2(x_i), \ldots, \sum_{i=1}^n t_k(x_i))$ is complete, as long as the parameter space Θ contains an open set in $\mathbb{R}^k$

NOTE:

field: 279

field: Suppose that a statistic T is complete and let g be a one-to-one function. Is the statistic $U = g(T)$ also complete?

field: Yes

NOTE:

field: 280

field: Does complete statistic imply minimial sufficient statistic?

field: If a minimal sufficient statistic exists, then any complete statistic is also a minimal sufficient statistic

NOTE:

field: 281

field: Basu's Theorem

field: If $T(\mathbf{x})$ is a complete and minimal sufficient statistic, then $T(\mathbf{x})$ is an independent of every ancillary statistic.

NOTE:

field: 282

field: Likelihood function

field: Let $f(\mathbf{x}|\theta)$ denote the joint pdf or pmf of the sample $\mathbf{X} = (X_1, \dots, X_n)$, then given that $\mathbf{X} = \mathbf{x}$ is observed, the function of θ defined as

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$$

is called the Likelihood Function

NOTE:

field: 283

field: Idea of likelihood function

field: Suppose that $\mathbf{X}$ is a discrete random vector (so we can interpret probabilities easier)

Then $L(\theta|\mathbf{x}) = P_\theta(\mathbf{X} = \mathbf{x})$. Now if we compare the likelihood function at two parameter values θ_1, θ_2 and we observe that

$$P_{\theta_1}(\mathbf{X} = \mathbf{x}) = L(\theta_1|\mathbf{x}) > L(\theta_2|\mathbf{x}) = P_{\theta_2}(\mathbf{X} = \mathbf{x}$$

Then, the sample point $\mathbf{x}$ that we actually observed is more likely to have occured if $\theta = \theta_1$, than if $\theta = \theta_2$, which can be interpreted as that θ_1, is a more plausible value for the true value of θ than θ_2 is.

NOTE:

field: 284

field: Fisher information - one parameter case

field: Let X be a random variable with pdf or pmf $f(x|\theta)$ where $\theta \in \Theta \subseteq \mathbb{R}$ (Fisher) information about θ contained in X is

$$I_X(\theta) = E_\theta[\big(\frac{\partial}{\partial\theta} \log f(x|\theta)\big)^2]$$

NOTE:

field: 285

field: Example of one parameter case Fisher information Suppose that $X \sim Bern(p)$ What is the information that X contains about the parameter p?

field: We have that $f(x|p) = p^x(1-p)^{1-x}$. Then

$$\log f(x|p) = x\log p + (1-x)\log(1-p)$$

$$\frac{\partial}{\partial p}\log f(x|p) = \frac{x}{p} - \frac{1-x}{1-p}$$

We obtain

$$\begin{aligned}\big(\frac{\partial}{\partial p}\log f(x|p)\big)^2 &= \big(\frac{x}{p} - \frac{1-x}{1-p}\big)^2 \\ &= \frac{x^2}{p^2} - \frac{2x(1-x)}{p(1-p)} + \frac{(1-x)^2}{(1-p)^2} \\ &= \frac{x^2}{p^2} - \frac{2(x-x^2)}{p(1-p)} + \frac{(1-2x+x^2)}{(1-p)^2}\end{aligned}$$

Therefore,

$$\begin{aligned}I_x(p) &= E_p[(\frac{\partial}{\partial p}\log f(x|p))^2] \\ &= \frac{p}{p^2} - \frac{2(p-p)}{p(1-p)} + \frac{1-2p+p}{(1-p)^2} \\ &= \frac{1}{p} + \frac{1}{1-p} \\ &= \frac{1}{p(1-p)}\end{aligned}$$

NOTE:

field: 286

field:

$$I_x(\theta) = E_\theta\big[(\frac{\partial}{\partial\theta}\log f(x|\theta))^2\big] =$$

field: If $f(x|\theta)$ satisfies

$$\frac{\partial}{\partial\theta}E_\theta\big(\frac{\partial}{\partial\theta}\log f(x|\theta)\big) = \int \frac{\partial}{\partial\theta}\big[\frac{\partial}{\partial\theta}\log f(x|\theta)\big]f(x|\theta)dx$$

$$I_x(\theta) = E_\theta\big[(\frac{\partial}{\partial\theta}\log f(x|\theta))^2\big] = -E_\theta(\frac{\partial^2}{\partial\theta^2}\log f(x|\theta))$$

NOTE:

field: 287

field: Suppose that $X_1,\ldots,X_n$ are iid observations with common pdf or pmf $f(x|\theta)$. Then, the information about θ contained in the sample $\mathbf{X} = (X_1,\ldots,X_n)$ is

field:

$$I_{\mathbf{X}}(\theta) = nI_{X_1}(\theta)$$

NOTE:

field: 288

field: Fisher Information - multiparameter case

field: Let X be a random variable with pdf or pmf $f(x|\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1,\theta_2) \in \Theta \subseteq \mathbb{R}^2$. Denote by

$$I_{ij}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\big[(\frac{\partial}{\partial\theta_i}\log f(x|\boldsymbol{\theta}))(\frac{\partial}{\partial\theta_j}\log f(x|\boldsymbol{\theta}))\big] = -E_{\boldsymbol{\theta}}[\frac{\partial}{\partial\theta_i\theta_j}\log f(x|\boldsymbol{\theta})]$$

For $i,j = 1,2$. Then the (fisher) information matrix about $\boldsymbol{\theta}$ is

$$I_x(\boldsymbol{\theta}) = \begin{pmatrix} I_{11}(\boldsymbol{\theta}) & I_{12}(\boldsymbol{\theta}) \\ I_{21}(\boldsymbol{\theta}) & I_{12}(\boldsymbol{\theta}) \end{pmatrix}$$

NOTE:

field: 289

field: Find Fisher information for Normal RVs

field: We have that $\boldsymbol{\theta} = (\mu, \sigma^2)$ and $f(x|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$
Then,

$$\frac{\partial}{\partial \mu} \log f(x|\boldsymbol{\theta}) = \frac{\partial}{\partial}[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x-\mu)^2] = \frac{(x-\mu)}{\sigma^2}$$

$$\frac{\partial}{\partial \sigma^2} = \frac{1}{2\sigma^2}[\frac{(x-\mu)^2}{\sigma^2} - 1]$$

Therefore $I_{11} = E_\theta[(\frac{\partial}{\partial \mu}\log f(x|\boldsymbol{\theta}))^2] = E_\theta[\frac{(x-\mu)^2}{\sigma^4}] = \frac{1}{\sigma^4}\sigma^2 = \frac{1}{\sigma^2}$

$$\begin{aligned}
I_{22}(\boldsymbol{\theta}) &= E_\theta[\frac{\partial}{\partial \sigma^2}\log f(x|\boldsymbol{\theta})^2] \\
&= E_\theta\{[\frac{1}{2\sigma^2}(\frac{(x-\mu)^2}{\sigma^2} - 1)]^2\} \\
&= \frac{1}{4\sigma^4}E_\theta[(\frac{(x-\mu)^2}{\sigma^2} - 1)^2] \\
&= \frac{1}{4\sigma^4 \cdot 2} \\
&= \frac{1}{2\sigma^4} \quad \text{Since } = V(\chi^2_1)
\end{aligned}$$

Now for the off diagonal elements,

$$\begin{aligned}
I_{12}(\boldsymbol{\theta}) = I_{22}(\boldsymbol{\theta}) &= E_\theta\big[(\frac{\partial}{\partial \mu}\log f(x|\theta)(\frac{\partial}{\partial \sigma^2}\log f(x|\theta)))\big] \\
&= E_\theta\big[\frac{(x-\mu)}{\sigma^2}\frac{1}{2\sigma^2}[\frac{x-\mu}{\sigma^2} \cdot 1]\big] \\
&= \frac{1}{2\sigma^4}E_\theta[\frac{(x-\mu)^3}{\sigma^3} - (x-\mu)]
\end{aligned}$$

But $E_\theta[(x-\mu)^3] = E_\theta[(x-\mu)] = 0$, because X is symmetric around μ, and we obtain $I_{12}(\boldsymbol{\theta}) = I_{21}(\boldsymbol{\theta}) = 0$

We obtain that

$$
\begin{aligned}
I_{x_1}(\boldsymbol{\theta}) &= \begin{pmatrix} I_{11}(\boldsymbol{\theta}) & I_{12}(\boldsymbol{\theta}) \\ I_{21}(\boldsymbol{\theta}) & I_{22}(\boldsymbol{\theta}) \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}
\end{aligned}
$$

And hence

$$
I_{\mathbf{x}}(\boldsymbol{\theta}) = nI_{X_1}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}
$$

NOTE:

field: 290

field: $I_T(\theta) \leq$

field: $I_T(\theta) \leq I_{\mathbf{X}}(\theta)$ (The information of the statistic is less than or equal to the information of the sample)

NOTE:

field: 291

field: Let $\mathbf{X} = X_1, \ldots, X_n$ denote the entire data, and let $T = T(\mathbf{X})$ be some statistic. Then, for all $\theta \in \Theta \subseteq \mathbb{R}$, $I_{\mathbf{X}}(\theta) \geq I_t(\theta)$ Where the equality is attained...

field: Let $\mathbf{X} = X_1, \ldots, X_n$ denote the entire data, and let $T = T(\mathbf{X})$ be some statistic. Then, for all $\theta \in \Theta \subseteq \mathbb{R}$, $I_{\mathbf{X}}(\theta) \geq I_t(\theta)$ Where the equality is attained if and only iff $T(\mathbf{X})$ is sufficient for θ

NOTE:

field: 292

field: Let $\mathbf{X} = (X_1, \ldots, X_n)$, denote a sample of iid observations and suppose the statistic $T(\mathbf{X}) = (T_1(\mathbf{X}), T_2(\mathbf{X}))$ is such that T_1 and T_2 are independent. Then

$$I_T(\boldsymbol{\theta}) =$$

field: Let $\mathbf{X} = (X_1, \ldots, X_n)$, denote a sample of iid observations and suppose the statistic $T(\mathbf{X}) = (T_1(\mathbf{X}), T_2(\mathbf{X}))$ is such that T_1 and T_2 are independent. Then

$$I_T(\boldsymbol{\theta}) = I_{T_1}(\boldsymbol{\theta}) + I_{T_2}(\boldsymbol{\theta})$$

NOTE:

field: 293

field: Point estimator

field: Any statistic $T(\mathbf{X})$ that is used to estimate the value of a parameter is called a point estimator of θ. We write $\hat{\theta} = T(\mathbf{X})$

NOTE:

field: 294

field: Method of moments

field:

$$
\begin{aligned}
m_1 &= \frac{1}{n}\sum_{i=1}^{n} X_i^1, \quad \mu_1 = E(X^1) \\
m_2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2, \quad \mu_2 = E(X^2) \\
&\vdots \\
m_k &= \frac{1}{n}\sum_{i=1}^{n} X_i^k \quad \mu_k = E(X^k)
\end{aligned}
$$

Equating and solving for θ gives the MoM estimators

NOTE:

field: 295

field: Example Method of Moments Suppose that $X_1, \ldots, X_n$ are iid Binomial(k, p), where both k and p are unknown.

field: We have that

$$
P(X_i = x|k, p) = \binom{k}{x} p^x (1-p)^{k-x}, x = 0, 1, \ldots, k
$$

and we obtain $E(X_1) = kp$, $E(X_1^2) = kp(1-p) + k^2p^2$

Solving the sytem of equations we obtain

$$
\begin{aligned}
m_1 &= \frac{1}{n}\sum_{i=1}^{n} X_i = kp \\
m_2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 = kp(1-p) + k^2p^2
\end{aligned}
$$

Sovling the system of equations:

$$\tilde{p} = \frac{\bar{x}}{\tilde{k}}$$
$$\tilde{k} = \frac{\bar{x}^2}{\bar{x} - \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Possible problems: k has to be an integer, and not negative. (Estimates of parameters that are outside of the parameter space.)

NOTE:

field: 296

field: Maximum Likelihood Estimator

field: In this context, we define the **Maximum Likelihood Estimator (MLE)** of θ as the parameter value $\hat{\theta}_{ML} = \hat{\theta}(\mathbf{x})$ that satisfies

$$L(\hat{\theta}_{ML}|\mathbf{x}) = \sup_{\theta\in\Theta} L(\theta|\mathbf{x})$$

Note this often proceedes as taking the derivative of the log likelihood function and setting to zero to solve for parameters - not always

NOTE:

field: 297

field: Example of MLE Suppose that $X_1, \ldots, X_n$ are iid Exponential(λ). Find the MLE $\hat{\lambda}_{ML}$ of λ

field: Suppose that $X_1, \ldots, X_n$ are iid Exponential(λ). Find the MLE $\hat{\lambda}_{ML}$ of λ

We have that $f(x|\lambda) = \frac{1}{\lambda}e^{x/\lambda}$, $x > 0$, and therefore

$$L(\lambda|x) = \prod_{i=1}^{n} \frac{1}{\lambda} e^{x_i/\lambda} = \lambda^{-n} e^{-\frac{1}{\lambda}\sum_{i=1}^{n} x_i}$$

Since $\log(\cdot)$ is a strictly monotone (one-to-one) and increasing, we consider instead the maximization of the log-likelihood

$$l(\lambda|\mathbf{x}) = \log L(\lambda|\mathbf{x}) = -n\log\lambda - \frac{1}{\lambda}\sum_{i=1}^{n} x_i$$

$$\frac{\partial}{\partial\lambda} l(\lambda|\mathbf{x}) = \frac{-n}{\lambda} + \frac{1}{\lambda^2}\sum_{i=1}^{n} x_i$$

Solving $\frac{\partial}{\partial\lambda} l(\lambda|\mathbf{x}) = 0$, we obtain

$$\frac{-n}{\lambda} + \frac{1}{\lambda^2}\sum_{i=1}^{n} x_i = 0$$

$$-n\lambda + n\bar{x} = 0$$

$$\lambda = \bar{x}$$

NOTE:

field: 298

field: Example of MLE when can't differemtiate

Suppose that $X_1, \ldots, X_n$ are iid Uniform$(0, \theta)$, $\theta > 0$. Find the MLE of θ

field: We have that $f(x|\theta) = \frac{1}{\theta} I(0 < x < \theta)$

And therefore

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\theta} I(0 < x_i < \theta)$$

$$= \frac{1}{\theta^n} I(X_{(1)} > 0) I(X_{(n)} < \theta)$$

In this case, the support of X depends on θ and the maximization problem only makes sense whenever $L(\theta|\mathbf{x}) > 0$. We cannot simply approach the problem by taking partial derivatives, but assuming the likelihood is positive, we notice that $L(\theta|\mathbf{x})$ is decerasing as a function of θ, for $\theta > X_{(n)}$

Picture with $L(\theta)$ as zero untill $X_{(n)}$ on x axis, goes up to $1/X_{(n)}$ there and decreases with $\frac{1}{\theta^n}$

It follows the MLE of θ is $\hat{\theta}_{ML} = X_{(n)}$

NOTE:

field: 299

field: If $\hat{\theta}_{ML}$ is the MLE of θ, then for any function $\tau(\theta)$, the MLE of $\eta = \tau(\theta)$ is $\hat{\eta}_{ML} =$

field: If $\hat{\theta}_{ML}$ is the MLE of θ, then for any function $\tau(\theta)$, the MLE of $\eta = \tau(\theta)$ is $\hat{\eta}_{ML} = \tau(\hat{\theta}_{ML})$

NOTE:

field: 300

field: Bias

field: Let $\hat{\theta} = T(\mathbf{X})$ be an estimatro of θ. Then the Bias of $\hat{\theta}$ as an estimator of θ is defined as

$$B_\theta(\hat{\theta}) = E_\theta(\hat{\theta} - \theta) = E_\theta(\hat{\theta}) - \theta$$

That is the difference between the expected value of $\hat{\theta}$ and θ.

An estimator $\hat{\theta}$ of θ is said to be unbiased if $B_\theta(\hat{\theta}) = 0 \quad \forall\theta$

NOTE:

field: 301

field: Mean Squared Error

field: Let $\hat{\theta} = T(\mathbf{X})$ be an estimate of θ. Then, the **Mean Squared Error** (MSE) of $\hat{\theta}$ as an estimator of θ is defined as:

$$MSE(\hat{\theta}) = E_\theta[(\hat{\theta} - \theta)^2] = V_\theta(\hat{\theta}) + [B_\theta(\hat{\theta})]^2$$

NOTE:

field: 302

field: Do unbiased estimators always exist?

field: No, Suppose that $X \sim \text{Binomial}(n, p)$ and let $\theta = 1/p$ be the parameter of interest. Can we find an unbiased estimator for θ?- No

NOTE:

field: 303

field: UMVUE

field: An estimator W^* is called a best unbiased estimator of $\tau(\theta)$ if it satisfies $E_\theta(W^*) = \tau(\theta)$, for all θ, and for any other estimator W with $E_\theta(W) = \tau(\theta)$, we have $V_\theta(W^*) \leq V_\theta(W), \forall\theta$. Equivalently W^* is also called a **Uniform Minimal Variance Unbiased Estimator** (UMVUE) of $\tau(\theta)$

NOTE:

field: 304

field: Finding a UMVUE

field: Start with a complete statistic, (find min suff statistic, prove completeness), Find bias (ie $E(T(\mathbf{X}))$). Then adjust $T(\mathbf{X})$ to be unbiased. (ie center or scale)

NOTE:

field: 305

field: Cramer-Rao Inequality

field: Let $X_1, \ldots, X_n$ be a sample with joint pdf or pmf $f(\mathbf{x}|\theta)$ and let $W(\mathbf{X}) = W(X_1, \ldots, X_n)$ be any estimator satisfying

$$\frac{d}{d\theta} E_\theta(W(X)) = \int \frac{d}{d\theta}[W(\mathbf{X}) f(\mathbf{x}|\theta)] d\mathbf{x}$$

and $V_\theta(W(\mathbf{X})) < \infty$

Then,

$$V_\theta(W(\mathbf{X})) \geq \frac{(\frac{d}{d\theta} E_\theta(W(\mathbf{X})))^2}{E_\theta[(\frac{\partial}{\partial\theta} \log f(\mathbf{x}|\theta))^2]}$$

Observe that if the sample $X_1, \ldots, X_n$ is iid with common pdf or pmf $f(x|\theta)$, we obtain

$$V_\theta(W(\mathbf{X})) \geq \frac{[\frac{d}{d\theta} E_\theta(W(\mathbf{X}))]^2}{n E_\theta[(\log f(\mathbf{x}|\theta))^2]}$$

The denominator is the information in the sample about θ

We have that as the information number gets bigger we have a smaller bound for the variance. of the best unbiased estimator and therefore more information is available.

NOTE:

field: 306

field: Cramer-Rao and UMVUE example UMVUE of λ for Poisson

field: Poisson example, we have $\tau(\lambda) = \lambda$, so $\frac{d}{d\lambda}\tau(\lambda) = 1$

On the other hand,

$$\begin{aligned}
nE_\lambda[(\frac{d}{d\lambda}\log f(x|\lambda))^2] &= -nE_\lambda(\frac{\partial^2}{\partial\lambda^2})\log f(x|\lambda)) \\
&= -nE_\lambda(\frac{\partial^2}{\partial\lambda^2}\log(\frac{e^{-\lambda}\lambda^x}{x!})) \\
&= -nE_\lambda[\frac{\partial^2}{\partial\lambda^2}(-\lambda + x\log\lambda - \log(x!))] \\
&= -nE_\lambda(\frac{-x}{\lambda^2}) \\
&= \frac{n}{\lambda}
\end{aligned}$$

Therefore, for any unbiased estimator W of λ, we must have $V_\lambda(W) \geq \lambda/n$. Since $V_\lambda(\bar{X}) = \frac{\lambda}{n}$, we have that $\bar{X}$ is an UMVUE of λ

NOTE:

field: 307

field: Does S^2 for Normal attain cramer rao?

field: No - Suppose that $X_1, \ldots, X_n$ are iid N(μ, σ^2) and consider the estimation of σ^2 when μ is unknown.

We have that

$$\frac{\partial^2}{\partial(\sigma^2)^2}\log[\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}] = \frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}$$

and

$$\begin{aligned}
-E[\frac{\partial^2}{\partial(\sigma^2)^2}\log f(x|\mu,\sigma^2)] &= -E(\frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}) \\
&= -\frac{1}{2\sigma^4} + \frac{\sigma^2}{\sigma^6} \\
&= \frac{1}{2\sigma^4}
\end{aligned}$$

and therefore, any unbiased estimator W of σ^2 must satisfy $V(W) \geq \frac{2\sigma^4}{n}$. Recall that for S^2 we have

$$V(S^2) = \frac{2\sigma^4}{n-1} > \frac{2\sigma^4}{n}$$

and therefore S^2 does not attain the cramer-rao lower bound.

NOTE:

field: 308

field: Rao-Blackwell

field: Let W be any unbiased estimator $\tau(\theta)$ and let T be a sufficient statistic for θ. Define $\phi(T) = E(W|T)$. Then $E_\theta(\phi(T)) = \tau(\theta)$ and $V_\theta(\phi(T)) \leq V_\theta(W)$, for all θ That is, $\phi(T)$ is a uniformly better unbiased estimator of $\tau(\theta)$

NOTE:

field: 309

field: Use of Rao-Blackwell

field: Estimators can be improved (their MSE) using sufficiency (already sufficient statistics, or functions of sufficient statistics cannot be improved)

NOTE:

field: 310

field: Are unbiased estimators based on complete sufficient statistics unique.

field: Unbiased estimators based on complete sufficient statistics are unique.

tags: Theory3

NOTE:

field: 311

field: Data summaries vs Prediciton vs Inference

field:

- Data summaries: descriptive statistics summarizing a dataset (ie sample mean)
- Prediction: Use patterns in a data-set to make predictions regarding values of new observations
  - Prediction setting is more flexible than inference setting, as we are not trying to make probabilistic inference, assumptions only matter if they affect prediction quality.
- Inference: Use observations in data set to infer information concerning population parameters

NOTE:

field: 312

field: Parametric Inference

field: Inference (estimation and/or hypothesis testing performed under the assumption taht the data come from a popluation distribution that belongs to some family of distributions $F(x;\theta)$) parametrized by a finite-dimensional parameter θ

Parameter space: The set Θ of all possible values of the parameter θ

Vs Nonparametric Inference - where no or limited assumptions or specifications of the form of the population distributions

NOTE:

field: 313

field: Are the following tests parametric, semiparametric, or nonparametric

- F-test
- Exact binomial test
- Fisher's exact test
- t-test
- Wilcoxon rank sum
- Permutation tests
- Sign test
- Mood's test
- KS test
- t-test

field:

- F-test - Parametric
- Exact binomial test
- Fisher's exact test
- t-test
- Wilcoxon rank sum: semiparametric
- Permutation tests
- Sign test: nonparametric
- Mood's test
- KS test
- t-test

NOTE:

field: 314

field: Definition: Simple hypothesis, composite hypothesis

field:

- Simple hypothesis: Completely specifies the parameter value and therefore the population distribution. Simple hypothesis have the form $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$, for specified values of θ_0 and θ_1
- Composite hypothesis: Includes more than one possible parameter value. Composite hypotheses have the form $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_1$

NOTE:

field: 315

field: Test procedure:

field:

- Random Sample (data): $X_1, \cdots, X_n$
- Sample Space $\mathscr{X}$ the set of all possible observed samples $X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$
- Hypothesis $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_1$ with $\Theta_0 \cap \Theta_1 = \emptyset$
- Rejection Region $\mathscr{R} \subset \mathscr{X}$:
  - If $(X_1, \ldots, X_n \in \mathscr{R}$, Reject H_0)
  - If $X_1, \ldots, X_n \notin \mathscr{R}$, Fail to reject H_0

Equivalently

- Random Sample (data): $X_1, \cdots, X_n$

- Test statistic $T(X_1, \cdots, X_n)$ is some function of the data, which is itself a random variable
- Test Statistic Sample Space $\mathscr{T}$ the set of all possible observed samples $T = t$
- Rejection Region $\mathscr{R}_T \subset \mathscr{T}$:
  - If $T(X_1, \ldots, X_n) \in \mathscr{R}_t$, Reject H_0)
  - If $T(X_1, \ldots, X_n) \notin \mathscr{R}_t$, Fail to reject H_0

NOTE:

field: 316

field: Power function (definition)

field: We can summarize the performance of a test procedure through the power function:

$$\begin{aligned}
\text{Power}(\theta) = \beta(\theta) &= P_\theta(\text{Reject } H_0 \text{ when } \theta \text{is the true value of the parameter of interest}) \\
&= P_\theta((X_1, \ldots X_n) \in \mathscr{R}) \\
&= P_\theta(T(X_1, \ldots, X_n) \in \mathscr{R}_t)
\end{aligned}$$

Equivalently, for a critical function ψ,

$$\beta(\theta) = E_\theta(\psi(x_1, \ldots, x_n))$$

NOTE:

field: 317

field: Calculating Type I and Type II errors from the power function

field:

$$P(\text{Type I Error when } \theta = \theta_0 \in \Theta_0) = \beta(\theta_0)$$

$$P(\text{Type II Error when } \theta = \theta_1 \in \Theta_1) = 1 - \beta(\theta_1)$$

(Note these are for simple hypotheses), for complex hypothesis, we want to look a the maximum possible error

To work out these probabilities, we need to know the distribution of the test statistics under the null (For type I error) and alternative (For type II error)

NOTE:

field: 318

field: Size of a test procedure

field: The size of a test procedure for a null hypothesis $H_0 : \theta \in \Theta_0$ is the value

$$sup_{\theta \in \Theta_0} P_\theta(\text{Reject } H_0) = sup_{\theta \in \Theta_0} \beta(\theta)$$

That is, the size of a test procedure is the largest value of the probability of a Type I Error, across all values of θ in the null hypothesis set Θ_0

NOTE:

field: 319

field: Definition of a level α test

field: A hypothesis test procedure is said to be a level α test if

$$sup_{\theta \in \Theta_0} P_\theta(\text{Reject } H_0) = sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$$

That is if the size of the test is less than or equal to α, the test is a level α test.

NOTE:

field: 320

field: most powerful level α test (definition)

field: Let $\mathscr{C}_\alpha$ be the set of all tests of $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$ where $\Theta_0 \cap \Theta_1 = \emptyset$ that have level α. A test belonging to $\mathscr{C}_\alpha$ is the most powerful level α test at $\theta_1 \in \Theta_1$ if

$$\beta(\theta_1) \geq \beta^*(\theta_1)$$

for any other test in $\mathscr{C}_\alpha$ with power function $\beta^*(\theta)$

NOTE:

field: 321

field: Uniformly most powerful level α test (definition)

field: A test belonging to $\mathscr{C}_\alpha$ with power function $\beta(\theta)$ is uniformly most powerful level α if it is the most powerful for every $\theta_1 \in \Theta_1$

NOTE:

field: 322

field: Critical Function / Test Function (definition)

field: A function $\psi : \mathscr{X} \to [0, 1]$ such that $\psi(x_1, \ldots x_n)$ is the probability of rejecting H_0 when the sample $(X_1 = x_1, \ldots X_n = x_n)$ is observed is called a critical function of a test procedure.

NOTE:

field: 323

field: Randomized Test (definition)

field: A test procedure with critical function ψ for which there are some points in the sample space such that $0 < \psi < 1$ is called a randomized test (often used in discrete cases)

NOTE:

field: 324

field: Finding the most powerful level α test of a simple null hypothesis vs a simple alternative hypothesis

field: $(Neyman - Pearson)$ The most powerful level α test of a simple null hypothesis H_0 vs a simple alternative hypothesis H_1 based on data $\mathbf{X}$ is given by the critical function

$$\psi(\mathbf{X}) = \begin{cases} 1 & \text{if } \frac{L(H_0:x)}{L(H_1:x)} < k \\ c & \text{if } \frac{L(H_0:x)}{L(H_1:x)} = k \\ 0 & \text{if } \frac{L(H_0:x)}{L(H_1:x)} > k \end{cases}$$

Where the constants k and c are chosen to ensure that $E_{H_0}(\phi(\mathbf{X})) = \alpha$

NOTE:

field: 325

field: Steps for using Neyman-Pearson to obtain most powerful tests for simple and alternative hypotheses:

field:

1. Identify the likelihood under the null $L(H_0 : x)$ and alternative $L(H_0 : x)$
2. Determine how the ratio of the likelihoods $\frac{L(H_0:x)}{L(H_1:x)}$ depends on the observed data $\mathbf{x}$ (ie is it an increasing or decreasing function of $T(\mathbf{X})$)?
3. Identify the null distribution of the statistic $T(\mathbf{X})$

(a) If $\frac{L(H_0 : x)}{L(H_1 : x)}$ is an increasing function of $T(\mathbf{x})$, rejecting for small values of $\frac{L(H_0 : x)}{L(H_1 : x)}$ is equivalent to rejecting for small values of $T(\mathbf{x})$, so find k such that

$$P_{H_0}(T(\mathbf{x}) < k) = \alpha$$

(b) If $\frac{L(H_0 : x)}{L(H_1 : x)}$ is a decreasing function of $T(\mathbf{x})$, rejecting for large values of $\frac{L(H_0 : x)}{L(H_1 : x)}$ is equivalent to rejecting for large values of $T(\mathbf{x})$, so find k such that

$$P_{H_0}(T(\mathbf{x}) > k) = \alpha$$

NOTE:

field: 326

field: Uniformly most powerful (UMP) level α test procedure

field: Uniformly most powerful (UMP) level α test procedure for testing $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$ is one with power funciton $\beta(\theta)$ such that for every $\theta_1 \in \Theta_1$ we have

$$\beta(\theta) \geq \beta^*(\theta)$$

for any other level α test prodedure with power function $\beta^*(\theta)$

NOTE:

field: 327

field: Monotone likelihood ratio

field: The family of distributions $\{F(x|theta)\}$ indexed by parmeter $\theta \in \Theta$ has monotone likelihood ratio if there is a statistic $T(\mathbf{X})$ such that for all $\theta^* > \theta \in \Theta$ and $\mathbf{x} \in \mathscr{X}$, the likelihood ratio

$$\frac{L(\theta^*|\mathbf{x})}{L(\theta|\mathbf{x})} \quad \text{is monotone nondecreasing in } T(\mathbf{x})$$

NOTE:

field: 328

field: How to find the UMP test of a simple null hypothesis vs a one sided complex alternative

field: See if the family has monotone likelihood ratio in $T(\mathbf{x})$ UMP tests of one sided alternative hypothesis exist and are given by the form in Neyman-Pearson (by Karlin Rubin)

NOTE:

field: 329

field: Karlin-Rubin Theorem

field: Suppose the family of distributions $\{F(x|theta)\}$ indexed by parmeter $\theta \in \Theta$ has monotone likelihood ratio Then to test

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$$

the test funciton

$$\phi(\mathbf{X}) = \begin{cases} 1 & ifT(\mathbf{X}) > k \\ \gamma & ifT(\mathbf{X}) = k \\ 0 & ifT(\mathbf{X}) < k \end{cases}$$

Where k and γ are chosen so that $E_{\theta_0}(\phi(\mathbf{X})) = \alpha$ gives a uniformly most powerful (UMP) level α test.

(note if we have a one sided lower alternative, we flip the direction of the inequalities)

NOTE:

field: 330

field: Is there a UMP two sided test for $X_1, \ldots, X_n$ iid Exp(λ) where $H_0 : \lambda = 2$ vs $H_1 \lambda \neq 2$?

field: No: For $\lambda_1 > \lambda_0 = 2$,the UMP test would have the form

$$\psi(\mathbf{X}) = \begin{cases} 1 & \text{if } T(\mathbf{x}) = \sum_{i=1}^n x_i > k_1 \\ 0 & \text{if } T(\mathbf{x}) = \sum_{i=1}^n x_i < k_1 \end{cases}$$

and for $\lambda_1 < \lambda_0 = 2$, the UMP test would have the form

$$\psi(\mathbf{X}) = \begin{cases} 1 & \text{if } T(\mathbf{x}) = \sum_{i=1}^n x_i < k_2 \\ 0 & \text{if } T(\mathbf{x}) = \sum_{i=1}^n x_i > k_2 \end{cases}$$

Since these forms are not the same, there is no UMP test.

NOTE:

field: 331

field: Let $X \sim \text{Unif}(0, \theta)$. Is there a UMP test for testing two sided $H_0 : \theta = 1$ vs $H_1 : \theta \neq 1$

field: Yes:

$$\psi(\mathbf{x}) = \begin{cases} 1 & x < \alpha \text{ or } x > 1 \\ 0 & \alpha < x < 1 \end{cases}$$

NOTE:

field: 332

field: Unbiased test (definition)

field: A test of $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$ is called unbiased if $\beta(\theta_1) \geq \beta(\theta_0)$ for all $\theta_1 \in \Theta_1$ and all $\theta_0 \in \Theta_0$

NOTE:

field: 333

field: Uniformly most powerful unbiased (UMPU) level α test (definition)

field: A level α test of $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$ with critical function $\psi(\mathbf{x})$ is called uniformly most powerful unbiased (UMPU) if it is unbiased level α and for any other unbiased test with critical function $\psi^*(\mathbf{x})$, we have

$$E_\theta(\psi(\mathbf{x})) \geq E_\theta(\psi^*(\mathbf{x})) \quad \text{for all } \theta \in \Theta_1$$

NOTE:

field: 334

field: Likelihood Ratio Test (definition)

field: Suppose we have the data $\mathbf{X} = X_1, \ldots, X_n$, with joint density $f(x;\theta)$ for some parameter $\theta \in \Theta$, and we wish to perform a level α test of $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$, where $\Theta_1 \cup \Theta_0 = \Theta$. The likelihood ratio test statistic is given by

$$\lambda(\mathbf{x}) = \frac{sup_{\theta \in \Theta_0} L(\theta|x)}{sup_{\theta \in \Theta} L(\theta|x)} = \frac{L(\hat{\theta}_{0,MLE}; x)}{L(\hat{\theta}_{MLE}; x)}$$

and the null hypothesis is rejected for small values of λ (indicating that the null hypothesis is relatively 'unlikely')

We maximize by finding $\theta = \hat{\theta}_{MLE}$ and $\hat{\theta}_{0,MLE}$

NOTE:

field: 335

field: If $T(\mathbf{X})$ is a sufficient statistic for θ, then $\lambda(\mathbf{x})$ (the LRT statistic)...

field: If $T(\mathbf{X})$ is a sufficient statistic for θ, then $\lambda(\mathbf{x})$ will be a function of $T(\mathbf{x})$. In particular $\lambda(\mathbf{x})$ will be a function of the minimal sufficient statistic

NOTE:

field: 336

field: Frequentist Probability vs Bayesian probability (definition)

field:

- Frequentist: For an event E, in our outcome space, $P(E)$ is the long run proportion of experiments that have outcome E, the relative frequency with which an event happens is its probability
- Bayesian: For an event E in the outcome space, $P(E)$ is any number between zero and one that you want to assign it, as long as you are coherent about the rules of additivity etc.

NOTE:

field: 337

field: Treatment of population parameters, frequentist vs bayesian

field:

- Frequentist: A population parameter θ is some fixed (though generally unknown value) that belongs to some set of possible values Θ
- Bayesian: A population parameter θ is a random quantity that has a prior distribution

NOTE:

field: 338

field: Likelihood function (bayesian)

field: Given some value of the parameter θ, the distribution of the data $\mathbf{x}$ is $f(\mathbf{x};\theta)$ is the likelihood (a function of both the value θ and the data $\mathbf{x}$).

NOTE:

field: 339

field: Posterior Distribution (definition)

field: The posterior distribution of theta given the observed data $\mathbf{x}$ is

$$k(\theta; \mathbf{x}) = \frac{f(\mathbf{x}; \theta)h(\theta)}{\int_\theta f(\mathbf{x}; \theta)h(\theta)d\theta}$$

Posterior probability $\propto$ Likelihood $\times$ prior probability
Note that the posterior distribution is proportional to the numerator.

NOTE:

field: 340

field: Conjugate Priors

field: If the prior $h(\theta)$ belongs to some (parametric) family of distributions $\mathscr{P}$ and the likelihood $L(\theta; \mathbf{x})$ (the joint density of the data for any particular value of θ) is such that the posterior $k(\theta; \mathbf{x})$ belongs to the same family $\mathscr{P}$, then this family of priors is said to be conjugate for the likelihood $L(\theta; \mathbf{x})$ (ie the posterior family is the prior family if we choose a conjugate prior.)

NOTE:

field: 341

field: Non-informative Priors

field: A non-informative prior is intended to give as little information as possible about the value of the parameter of interest θ.

NOTE:

field: 342

field: Improper prior

field: an improper prior is a prior that does not integrate to one

NOTE:

field: 343 ©

field: Bayes Estimator

field: A Bayes estimator (with respect to the particular prior/likelihood) is the estimator that minimizes the Bayesian Risk

$$\delta^* = arginf_{\delta \in D} \int_{\Theta} R(\theta, \delta) h(\theta) d\theta$$

Where D is the set of all possible estimators for θ

The Bayes Estimator equivalently minimizes the posterior risk, given the observed data.

For squared-error loss, the Bayes estimate is the mean of the posterior distribution $k\theta(\mathbf{x})$:

$$\delta^* = \int_{\Theta} \theta k(\theta|\mathbf{x}) d\theta$$

NOTE:

field: 344

field: Maximum A Posteriori (MAP)

field: A MAP test selects the hypothesis H_0 or H_1 that has the highest posterior probability. (Bayesian.)

NOTE:

field: 345

field: Definition of a p-value

field:

- For testing null hypothesis H_0 vs alternative hypothesis H_1, the p-value $p(\mathbf{x})$ corresponding to the observed data, is the smallest value α for which H_0 would be rejected by a size α test
- Let $W(\mathbf{X})$ be a test statistic such that large values of W are evidence that H_1 is true, and therefore the null hypothesis H_0 is rejected for large $W(\mathbf{X})$ then a p-value can be defined as

  $$p(\mathbf{x}) = sup_{\theta \in \Theta} P_\theta(W(\mathbf{X}) \geq W(\mathbf{x}))$$

  This says that the p-value is the (largest in the null space) probability of obtaining a test statistic at least as extreme as the observed test statistic value.
- A p-value is just a function of the observed data; a test statistic

NOTE:

field: 346

field: Validity of p-value

field: A p-value is valid (exact) if for every $\theta \in \Theta_0$ and every value of $\alpha \in [0, 1]$, we have

$$P_\theta(p(\mathbf{X}) \leq \alpha) \leq \alpha$$

NOTE:

field: 347

field: Confidence interval (definition)

field: Suppose we have data $\mathbf{X}$ such that the (joint) density of our data give information about an unknown parameter θ. Then a $(1-\alpha)100$ confidence interval for θ is a random interval $[L(\mathbf{X}), U(\mathbf{X})]$ such that

$$inf_{\theta \in \Theta} P_\theta(L(\mathbf{X}) \leq \theta \leq U(\mathbf{X})) = 1 - \alpha$$

It is important to note that it is the limits of the interval $L(\mathbf{X}), U(\mathbf{X})$ that are the random quantities here.

NOTE:

field: 348

field: Construct a CI using a hypothesis test

field: A level $(1-\alpha)100$ confidence interval can be constructed by inverting a level α hypothesis test. This fact is known as the duality of confidence intervals and hypothesis testing. The confidence region $\mathscr{C}$ $\mathscr{C} = \{\theta_0 : H_0 : \theta = \theta_0$ would not be rejected at level $\alpha\}$

(ie solve for θ_0 to be in the center)

NOTE:

field: 349

field: Pivot

field: Suppose X comes from some parametric family $F(\mathbf{x} : \theta)$ indexed by parameter θ. A pivot, or pivotal quantity is a random variable $U = g(\mathbf{X}, \theta)$ that depends upon both the sample $\mathbf{X}$ and the unknown parameter θ for which the distribution of U does not depend on θ

NOTE:

field: 350

field: Finding a confidence interval for θ using the pivotal method

field:

1. Identify a pivotal quantity U and its distribution $F_U(u)$
2. Find a and b such that

$$P(a < U < b) = 1 - \alpha$$

   Let $F_U(u)$ denote the cdf of the pivot U, so then we can set

$$a = F_U^{-1}(c\alpha)$$
$$b = F_U^{-1}(1 - (1 - c)\alpha)$$

   For any $c \in [0, 1]$ (usually .5 to split up area on the tails evenly)

3. Solve the inequality $a < U < b$ for θ in the middle.

NOTE:

field: 351

field: Pivotal CI example: Let $Y \sim exp(\theta)$.

field:

1. Let $U = Y/\theta$, so $U \sim Exp(1)$. which doesn't depend on θ, so U is a pivotal quantity.
2. We must find a,b, such that $P(a \leq U \leq b) = 1 - \alpha$. We then solve $P(U \leq a) = \alpha/2$ and $P(b \leq U) = \alpha/2$ Solve for θ

$$P(a \leq U \leq b) = P(Y/b \leq \theta \leq Y/a)$$

NOTE:

field: 352

field: Finding a pivotal quantity

field:

- If θ is a location parameter, a possible pivot has the form $U = T(\mathbf{X}) - a(\theta)$
- If θ is a scale parameter, $U = T(\mathbf{X})/b(\theta)$ is a possible pivot
- If θ is a location-scale parameter, $U = (T(\mathbf{X}) - a(\theta))/b(\theta)$ is a possible pivot
- If neither, use how parameter is related to X, ie if $F_Y(y) = y^N$, use y^N as the pivot.

NOTE:

field: 353

field: Confidence Interval optimality criteria

field:

- Length
  - Length can be a function of sample size and critical value choice (Normal mean with known variance), other cases a random quantity (Normal mean with unknown variance)
  - When length is random, we typically want confidence intervals with shortest mean length.
  - Ie for normal variance with known mean, the expected length of the confidence interval is $E(L) = c\sigma^2 n$, so a shortest interval would be not using equal tails, but requires numerical computation.
- Convexity
  - In the case of a one-dimensional parameter, this means that the region should be a connected interval

- Agreement with a reasonable estimate and with a reasonable hypothesis test.
  - Example score test and wald CI
- Equal coverage probability for all θ

NOTE:

field: 354

field: $1-\alpha$ credible interval

field: (Bayesian) A $1-\alpha$ credible interval is an interval $[a,b]$ such that the credible probability of that interval is $1-\alpha$, that is

$$\int_{\theta=a}^{b} k(\theta|\mathbf{x})d\theta = 1-\alpha$$

Where the posterior distribution $k(\theta|\mathbf{x}) = \frac{f(\mathbf{x};\theta)h(\theta)}{\int_\theta f(\mathbf{x};\theta)h(\theta)} = \frac{p(\mathbf{x}|\theta)p(\theta)}{p(\mathbf{x})}$, where the posterior is proportional to the likelihood times the prior

In many cases, we can choose $c = .5$ and let $a =$ the $c \times \alpha$ quantile of the distribution, and $b =$ the $1-(1-c)\times\alpha$ quantile of the posterior distribution, but in cases when the posterior distribution is unimodal, we can obtain shorter intervals.

NOTE:

field: 355

field: Credible probability

field: (Bayesian) For any region $\mathscr{A} \subset \Theta$, the credible probability of the set $\mathscr{A}$ is

$$P(\theta \in \mathscr{A}|\mathbf{x}) = \int_{\theta\in\mathscr{A}} k(\theta|\mathbf{x})d\theta$$

NOTE:

field: 356

field: Highest posterior density $1-\alpha$ credible interval

field: For a unimodal posterior distribution $k(\theta|\mathbf{x})$ Highest posterior density $1-\alpha$ credible interval is the interval $[a,b]$ such that

1. $\int_{\theta=a}^{b} k(\theta|\mathbf{x})d\theta = 1-\alpha$
2. $k(a|\theta) = k(b|\theta)$
3. $a < \theta^* < b$ where θ^* is the mode of the posterior distribution.

This credible interval will be the shortest. Note it might not contain the Bayes estimator for θ if the posterior distribution is very highly skewed and has a heavy tail.

NOTE:

field: 357

field: Unimodal pdf

field: A pdf is unimodal if there exists some value x^* such that

1. For all $x \leq x^*, f(x)$ is non decreasing
2. For all $x \geq x^*, f(x)$ is non-increasing.

NOTE:

field: 358

field: Confidence Intervals for functions of parameters: Obtain a confidence interval for $\tau = g(\theta)$, given a confidence interval for θ

field:

1. Invert a test of $H_0 : \tau = \tau_0$ vs $H_1 : \tau \neq \tau_0$
2. Create a pivot $U = h(\mathbf{X}, \tau)$, and construct a pivotal interval for τ
3. Transform the CI for θ into an interval for τ
   - Coverage of this interval: $C_\tau = \{\tau_0 : \tau_0 = g(\theta_0) \text{ for } \theta_0 \in C_\theta\}$
   - NOT the same as $(g(L_\theta(\mathbf{X})), g(U_\theta(\mathbf{X})))$ (unless strictly monotone, otherwise coverage not the same.)

NOTE:

field: 359

field: Joint Confidence Intervals for Multivariate parameters

field:

1. Construct $(1 - \alpha^*)$ CIs for each component θ_j separately
2. Define $\mathscr{C} = \{\theta_0 : \theta_{0,j} \in \mathscr{C}_j \text{ for all } j = 1, \dots p\}$
3. Chose the univariate coverage levels $(1-\alpha^*)$ to ensure that the coverage level of the joint confidence region is at least the desired level $(1 - \alpha)$
4. Often choose $\alpha^* = \alpha/p$, although this often gives larger CIs - over-coverage
5. If parameters independent, if we set $1 - \alpha = (1 - \alpha^{**})^p$

NOTE:

field: 360

field: Example: Construct a CI for

$$X_1, \dots, X_n \sim iidExp(\mu_x, \sigma)$$

$$Y_1, \dots, Y_n \sim iidExp(\mu_x, \sigma)$$

field: $\kappa(\theta) = \mu_x - \mu_y$

- Sufficient statistic for parameter (μ_x, μ_y, σ) is

$$(X_{(1)}, Y_{(1)}, W = \sum_{i=1}^{n}(X_i - X_{(1)}) + \sum_{i=1}^{n}(Y_i - Y_{(1)})$$

- $X_{(1)} \sim exp(\mu_x, \sigma/n)$
- $Y_{(1)} \sim exp(\mu_y, \sigma/n)$
- $\frac{2W}{\sigma} \sim \chi^2_{4n-4}$
- Construct pivot: $U = \frac{|n(X_{(1)} - Y_{(1)} - \mu_x - \mu_y)}{W/(2n-2)} \sim F_{2,4n-4}$

NOTE:

field: 361

field: Determinant of a square matrix

field: The determinant of a square matrix $\mathbf{A}$ is denoted by $|\mathbf{A}|$ and is defined recursively

- For a (1×1) matrix $|\mathbf{A}| = a_{11}$
- For a $(p \times p)$ matrix $|\mathbf{A}| = \sum_{j=1}^{p} a_{ij}(-1)^{i+j}|\mathbf{A}_{-i,-j}|$
- Where $\mathbf{A}_{-i,-j}$ is the matrix obtained by removing the ith column and jth row from $\mathbf{A}$

NOTE:

field: 362

field: Properties of the determinant

- $\det(\mathbf{I_p})$
- $\det(\mathbf{A}^{-1})$
- $\mathbf{A}$ is invertible iff
- $\det(AB)$
- $\det(\mathbf{A}^t)$
- $\det(\mathbf{XAX}^{-1}) =$
- Relationship to eigenvalues

field:

- $\det(\mathbf{I_p}) = 1$
- $\det(\mathbf{A}^{-1}) = (\det(\mathbf{A}))^{-1}$
- $\mathbf{A}$ is invertible iff $\det(\mathbf{A} \neq 0$
- $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$
- $\det(\mathbf{A}^t) = \det(\mathbf{A})$
- $\det(\mathbf{XAX}^{-1}) = \det(\mathbf{A})$
- The determinant of $\mathbf{A}$ is equal to the product of all eigen values of $\mathbf{A}$

NOTE:

field: 363

field: Properties of the Multivariate Normal Distribution

field:

- If $\mathbf{X}$ has a multivariate normal distribution, then each element has a marginal normal distribution
- Random variables $X_1, \ldots X_p$ with marginal normal distributions DO NOT necessarily have a multivariate normal joint distribution
- All subsets of elements of $\mathbf{X}$ have a multivariate normal distribution
- All linear combinations of the components of X are normally distributed
- $\mathbf{X} + c \sim MVN(\mu + c, \Sigma)$
- $\mathbf{AX} \sim MVN(\mathbf{A}\mu, \mathbf{A}\Sigma A^t)$, where each element of $\mathbf{AX}$ is a linear combination of the random vector $\mathbf{X}$
- $Cov(X_j, X_k) = \sigma_{jk} = 0$ iff and only if X_j, X_k independent
- $Z = \Sigma^{-1/2}(\vec{X} - \vec{\mu}) \sim MVN(\vec{0}, I_p)$
- $Z^t Z \sim \chi^2_p$

NOTE:

field: 364

field: Distribution of sample mean and variance from Multivariate normal

field:

- $\bar{\mathbf{X}} \sim MVN(\vec{\mu}, \frac{1}{n}\Sigma)$
- $\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^t$

NOTE:

field: 365

field: $X \sim \chi^2_{v_1}, Y \sim \chi^2_{v_2}, X + Y \sim$

field: $\chi^2_{v_1+v_2}$

NOTE:

field: 366

field: Definition of consistent sequence of estimators

field: Suppose we have random variables $X_1, X_2, \ldots$, such that the collection $\{X_1, \ldots X_n\}$ gives information about a parameter θ. A sequence of estimators $W_n = W_n(X_1, \ldots, X_n)$ is a consistent sequence of estimators of the parameter θ if FOR EVERY $\theta \in \Theta$,

$$W_n \vec{p} \theta$$

NOTE:

field: 367

field: Mean square error

field:

$$MSE(\hat{\theta}, \theta) = E[(\hat{\theta} - \theta)^2] = Bias^2 + V(\hat{\theta})$$

NOTE:

field: 368

field: If a sequence of estimators W_n for a parameter θ satisfies

- $\lim_{n\to\infty} V_\theta(W_n) \to 0$
- $\lim_{n\to\infty} E_\theta(W_n) \to \theta$
- Equivalently, if $MSE_\theta(W_n;\theta) \to 0$

for all $\theta \in \Theta$, then

field: W_n is a consistent sequence of estimators for the parameter θ

NOTE:

field: 369

field: How to prove consistency:

field: If a sequence of estimators W_n for a parameter θ satisfies

- $\lim_{n\to\infty} V_\theta(W_n) \to 0$
- $\lim_{n\to\infty} E_\theta(W_n) \to \theta$
- Equivalently, if $MSE_\theta(W_n;\theta) \to 0$

for all $\theta \in \Theta$, then W_n is consistent.

NOTE:

field: 370x

field: Definition of asympototic distribution of Estimators

field: Suppose that a sequence of random variables W_n satisfies

$$k_n(W_n - \theta)\vec{d}F$$

- k_n is the stabilizing constant,
- F is the asympototic distribution
- $\sigma^2(\theta)$ the asymptotic variance (which can't depend on n)

NOTE:

field: 371x

field: $X_i \sim Exp(\sigma)$, $\sum X_i \sim$

field: $\sum X_i \sim Gamma(n, \sigma)$

NOTE:

field: 372

field: Relationship between asymptotic variance and limiting variance

field: limiting variance $\geq$ asymptotic variance, where the limiting variance is defined as $\lim_{n\to\infty} V(Y_n)$

NOTE:

field: 373

field: Score function

field:

$$U(\theta) = \frac{\partial}{\partial\theta} l(\theta; \mathbf{x}) = \sum_{i=1}^{n} \frac{\frac{\partial}{\partial\theta} f(x_i; \theta)}{f(x_i; \theta)}$$

NOTE:

field: 374

field: Notation and test construction $X_1, \ldots, X_n$ iid

- Density/Mass Function
- Likelihood
- Log likelihood
- Maximum likelihood estimator
- Score
- Information

field:

- Density/Mass Function : $f(x;\theta)$
- Likelihood : $L(\theta;\mathbf{x}) = \prod_{i=1}^{n} f(x_i;\theta)$
- Log likelihood $l(\theta;\mathbf{x}) = \log L(\theta;\mathbf{x}) = \sum_{i=1}^{n} \log f(x_i;\theta)$
- Maximum likelihood estimator $\hat{\theta}_n$: $\frac{\partial}{\partial\theta} l(\theta;\mathbf{x})\Big|_{\theta=\hat{\theta_n}} = 0$ (except if support depends on parameter of interest)
- Score: $U(\theta) = \frac{\partial}{\partial\theta} l(\theta;\mathbf{x}) = \sum_{i=1}^{n} \frac{\frac{\partial}{\partial\theta} f(x_i;\theta)}{f(x_i;\theta)}$
- Information: $I_1(\theta) = E\bigg(-\frac{\partial^2}{\partial\theta^2} \log f(x;\theta)\bigg)$

NOTE:

field: 375

field: Important Distribution of Maximum Likelihood Estimators

field: Note these assumptions hold in exponential families, Binomial, Poisson, etc

- X_i iid with common density function
- Identifiability: for $\theta_1 \neq \theta_2$, $f(x;\theta_1) \neq f(x;\theta_2)$
- Common support (the one we need to verify) The set of possible values for X does not depend on the value of the parameter θ
- Open Parameter Space (ie cant include 0 or 1 in bernoulli case)

NOTE:

field: 376

field: Asymptotic Normality of MLE

field: Under the common assumptions, The MLE $\hat{\theta}_n$ for θ satisfies

$$\frac{\hat{\theta})n - \theta}{\sqrt{\frac{1}{nI_1(\theta)}}} \vec{d} N(0,1)$$

Equivalently, $\hat{\theta}_n \dot{\sim} N(\theta, \frac{1}{nI_1(\theta)})$ This implies that the MLE is asymptotically the UMVUE (as $E(\hat{\theta}_n \approx \theta)$ and $Var(\hat{\theta}_n) \approx \frac{1}{I_n(\theta)}$), the CRLB var of unbiased estimate of θ

NOTE:

field: 377

field: Mann-Wald Theorem (for MLE)

field: For any differentiable $g(\cdot)$ with non-zero first derivative: if $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \frac{1}{I_1(\theta)})$, then

$$\sqrt{n(g(\hat{\theta}_n) - g(\theta))} \xrightarrow{d} N(0, \frac{[g'(\theta)]^2}{I_1(\theta)})$$

NOTE:

field: 378

field: Wald Hypothesis Tests and Confidence intervals

field: Since θ is not known, we cannot find the value of $I(\theta)$ exactly, but we can use $\hat{I}(\theta) = I(\hat{\theta}_n) \xrightarrow{p} I(\theta)$, where $\hat{\theta}_n$ is the MLE

- Test statistic:
$$W(\theta_0) = \frac{\hat{\theta}_n - \theta_0}{\sqrt{\frac{1}{nI_1(\hat{\theta}_n)}}}$$
- Two sided alternative: Reject $H_0 : \theta = \theta_0$ vs $\theta \neq \theta_0$ if
$$|W(\theta_0)| > z_{\alpha/2}$$
- Confidence interval:
$$\left(\hat{\theta}_n - z_{\alpha/2}\sqrt{\frac{1}{nI_1(\hat{\theta_n})}}, \left(\hat{\theta}_n + z_{\alpha/2}\sqrt{\frac{1}{nI_1(\hat{\theta_n})}}\right)\right.$$

NOTE:

field: 379

field: Asymptotic Likelihood-Based Test Statistics

field:

Test	Test Statistic	Asymptotic Null Distribution
Likelihood Ratio	$G(\theta_0) = 2\big(l(\hat{\theta}_n; \mathbf{x}) - l(\theta_0; \mathbf{x})\big)$	χ^2_1
Wald	$W(\theta_0) = \frac{(\hat{\theta}_n - \theta_0)^2}{\frac{1}{nI(\hat{\theta}_n)}}$ $W^*(\theta_0) = \frac{\hat{\theta}_n - \theta_0}{\sqrt{\frac{1}{nI(\hat{\theta}_n)}}}$	χ^2_1
Score	$S(\theta_0) = \frac{U(\theta_0)^2}{nI(\theta_0)}$ (no need to find MLE) $S^*(\theta_0) = \frac{U(\theta_0)}{\sqrt{nI(\theta_0)}}$	χ^2_1 $N(0,1)$

NOTE:

field: 380

field: Comparison between Likelihood Ratio, Wald and Score tests and intervals

field:

1. Likelihood Ratio:
   - Requires computation of $sup_{\theta \in \Theta} L(\hat{\theta}; \mathbf{x})$ (likelihood of MLE) and $L(\theta_0; \mathbf{x})$ (likelihood of null hypothesis)
   - Generally hardest to invert to form confidence intervals
2. Wald:
   - Only requires computation of MLE and Information under MLE
   - Tends to differ from the Score and LR tests more than they do from each other
   - Easiest to invert to form confidence intervals
3. Score:
   - Only requires computation of Information under Null
   - Has best power of the three tests for alternatives close to null
   - Typically similar in form to Wald, except variance is estimated under null instead of using MLE

All three tests are asymptotically equivalent under the null hypothesis when n is large. If the alternative is true, they may differ substantially, no matter the value of n

NOTE:

field: 381

field: Variance Stabilizing Transformation

field: To potentially improve confidence interval coverage in settings where there is a mean-variance relationship, we can try to find a function $g(\cdot)$ such that

$$(g'(\theta))^2 V(\hat{\theta}) = 1$$

, That is the asymptotic variance of $g(\hat{\theta})$ does not depend on the value of θ. This function $g(\cdot)$ is called teh variance stabilizing transformation for the estimate $\hat{\theta}$