ESE 524: Detection and Estimation Theory

Recitation 1: Probability review

Washington University in St. Louis

Administrative Remarks

- ESE 520 (or a similar course in probability) is a prerequisite for this course.

- Students are expected to know the materials in this set of slides. The materials in these slides are covered in ESE 520, which is a pre-requisite for this class. These slides are intended to help refresh your memory.

Probability Definition

- Let Ω be the set of possible outcomes of an experiment and A the set of desired outcomes of the experiment

- Denote $|A|$ the cardinality of the set A. For a finite set of numbers, the cardinality is the number of elements in A

- Then, we define the probability of A as: $P(A) = \frac{|A|}{|\Omega|}$

Conditional Probabilities and Bayes' Formula

- What is the probability of getting a grade "A" in this class? vs. What is the probability of getting a grade "A" in this class, given that I attend all the recitations?

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

- If B_i partition the entire probability set
$\Rightarrow P(A) = \Sigma_i P(A|B_i) \cdot P(B_i)$

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

- If A_i partition the probability space: $P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\Sigma_j P(B|A_j) \cdot P(A_j)}$

Conditional Probabilities and Bayes' Formula

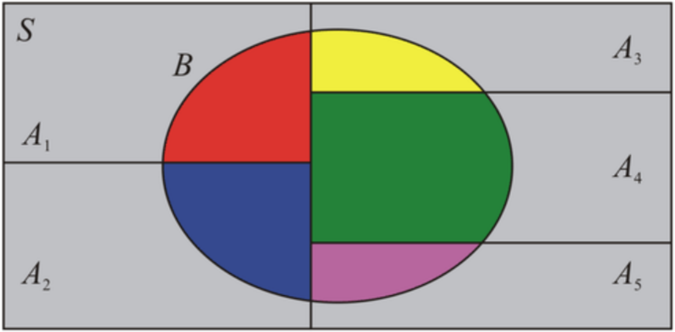
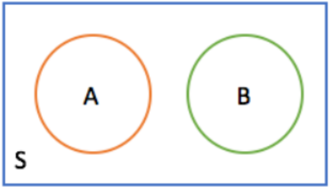


Figure 1: **Bayes theorem** Partition of the probability space.

Independent Events

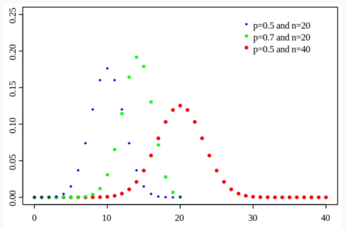
- A and B are independent event iff: $P(A \cap B) = P(A) \cdot P(B)$.

- Additionally, if A and B are independent: $P(A|B) = P(A)$



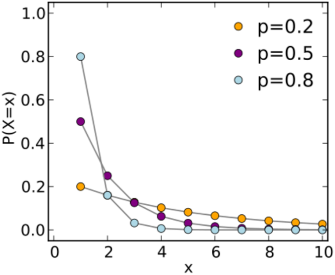
Examples of Discrete Probabilities

- The binomial distribution ($X \sim \mathrm{Bin}(n, p)$) models the probability of getting k successes in n experiments:
  $P_X(x = k) = \binom{n}{k}\, p^k q^{n-k}$



Examples of Discrete Probabilities

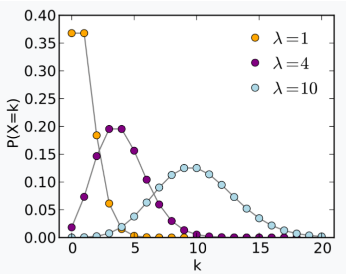
- The geometric distribution ($X \sim \mathrm{Geom}(p)$) models the probability of performing k experiments until we succeed once: $P_X(x = k) = p \cdot q^{k-1}$



Examples of Discrete Probabilities

- The Poisson Distribution $(X \sim \mathrm{Pois}(\lambda))$ models the probability of k events taking place in a given period of time:
  $P_X(x = k) = e^{-\lambda} \frac{\lambda^k}{k!}$

e.g., how many useful comments for the homework are given per hour of recitation:



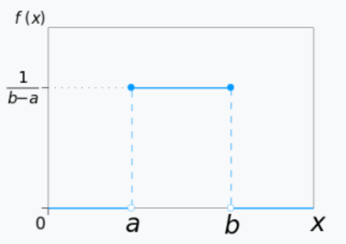
Continuous Probabilities

- Notation: X is a continuous random variable with probability distribution function $f_X(x)$.
- $P(a \leq x \leq b) = \int_a^b f_X(x)dx$
- $f_X(x) \geq 0$
- $\int_{-\infty}^{\infty} f_X(x)dx = 1$

Examples of Continuous Probabilities

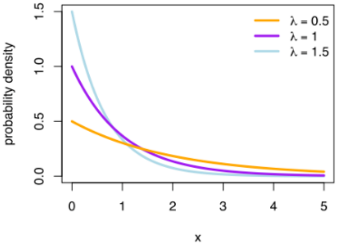
- The uniform distribution $X \sim \mathrm{U}(a, b)$ models an event that is equally likely to happen over a range of time:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$



Examples of Continuous Probabilities

- The exponential distribution $X \sim \exp(\lambda)$ models the time difference between events in a Poisson distribution:
  $f_X(x) = \lambda e^{-\lambda x}$

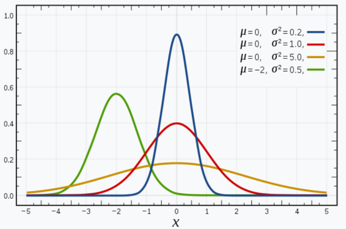


Examples of Continuous Probabilities

- The Normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$ is extremely useful as it models many experimental and statistical processes.

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

  Normalization: $z = \frac{x-\mu}{\sigma} \sim \mathcal{N}(0, 1)$



Cumulative Distribution Functions (c.d.f.)

- $F_X(x) \equiv P(X < x)$
- $F_X(x) = \int_{-\infty}^{x} f_X(x)dx$
- $F'_X(x) = f_X(x)$

Cumulative Distribution Functions (c.d.f.)

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

- $\lim_{x \to +\infty} F_X(x) = 1$

- $\lim_{x \to -\infty} F_X(x) = 0$

- $F_X(x)$ is a non-decreasing function

- $F_X(x)$ is a continuous function iff x is a continuous random variable

Transformation of Random Variables

Let x be a random variable with p.d.f. $f_x(x)$. Let $h(x)$ be a strictly monotonic and differentiable function and define $y = h(x)$, then:

$$f_Y(y) = \left|(h^{-1}(y))'\right| \cdot f_X(h^{-1}(y))$$

Transformation of Random Variables Example

- Given:
  - Random variable X with pdf $p_x(x) = x\exp(\frac{-x^2}{2})$
  - Standard Normal Random variable Y with pdf $p_y(y) = \frac{1}{\sqrt{2\pi}}\exp(\frac{-y^2}{2})$
  - constant c

  Find the joint distribution of $U = g_1(x, y) = \sqrt{X^2 + Y^2}$ and $V = g_2(x, y) = \frac{cY}{X}$.
- First solve for the inverse transformation (see board for details):

$$x = h_1(u, v) = \frac{u}{\sqrt{1 + \frac{v^2}{c^2}}}$$

$$y = h_2(u, v) = \frac{uv}{c\sqrt{1 + \frac{v^2}{c^2}}}$$

Transformation of Random Variables Example Continued

- Then find the Jacobian matrix:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{1+v^2/c^2}} & \frac{-uv}{c^2(1+v^2/c^2)^{3/2}} \\ \frac{v}{c\sqrt{1+v^2/c^2}} & \frac{u}{c(1+v^2/c^2)^{3/2}} \end{bmatrix}$$

- The determinant of $\mathbf{J}$ is $\det(\mathbf{J}) = \frac{cu}{c^2+v^2}$
- Following the transformation of variables formula:

$$p_{u,v}(u,v) = p_{x,y}(h_1(u,v), h_2(u,v))|\det(\mathbf{J})|$$

$$= \frac{u}{\sqrt{1+v^2/c^2}} \frac{e^{-u^2/2}}{\sqrt{2\pi}} \frac{cu}{c^2+v^2}$$

Transformation of Random Variables Example (Cont.)

- After some algebra, we can separate $p_{u,v}$ into:

$$p_{u,v} = \underbrace{\sqrt{\frac{2}{\pi}} u^2 e^{-u^2/2}}_{\text{Maxwell}} \cdot \underbrace{\frac{1}{2c}\left(1 + \frac{v^2}{c^2}\right)^{-3/2}}_{\text{t}}$$

  which is the product of a Maxwell distribution and student's-t distribution.

- Since $p_{u,v}(u,v) = p_u(u)p_v(v)$ these variables are independent.

Examples of Expected Value of Discrete Random Variables

- $\mathrm{E}X \equiv \sum_{\Omega} x P_X(x)$
- $X \sim \mathrm{Bin}(n, p) \Rightarrow \mathrm{E}X = np$
- $X \sim \mathrm{Geom}(p) \Rightarrow \mathrm{E}X = 1/p$
- $X \sim \mathrm{Pois}(\lambda) \Rightarrow \mathrm{E}X = \lambda$

Examples of Expected Value of Continuous Random Variables

- $\mathrm{E}X \equiv \int_{-\infty}^{x} x f_X(x) dx$
- $X \sim \mathrm{U}(a, b) \Rightarrow \mathrm{E}X = \frac{b-a}{2}$
- $X \sim \exp(\lambda) \Rightarrow \mathrm{E}X = 1/\lambda$
- $X \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow \mathrm{E}X = \mu$

Properties of the Expected Value

The expected value is a linear operator:

- $\mathrm{E}(cX + a) = c\mathrm{E}(X) + a$ (for a and c deterministic)
- $\mathrm{E}\left(g(X) + h(X)\right) = \mathrm{E}\left(g(X)\right) + \mathrm{E}\left(h(X)\right)$

- Lemma: If $f_X(x)$ is a symmetric function around a, then: $\mathrm{E}X = a$
- Law of the unconscious statistician: If $Y = h(x)$, then: $\mathrm{E}Y = \int_{-\infty}^{\infty} h(x) f_X(x) dx$

Variance

- $\mathrm{Var}(X) \equiv \mathrm{E}(X - \mathrm{E}X)^2$
- $X \sim \mathrm{Bin}(n, p) \Rightarrow \mathrm{var}(X) = npq$
- $X \sim \mathrm{Geom}(p) \Rightarrow \mathrm{var}(X) = q/p^2$
- $X \sim \mathrm{Pois}(\lambda) \Rightarrow \mathrm{var}(X) = \lambda$
- $X \sim \mathrm{U}(a, b) \Rightarrow \mathrm{var}(X) = \frac{(b-a)^2}{12}$
- $X \sim \exp(\lambda) \Rightarrow \mathrm{var}(X) = 1/\lambda^2$
- $X \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow \mathrm{var}(X) = \sigma^2$

Properties of the Variance

- $\mathrm{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}X)^2$

- $X = \mathrm{const} \Rightarrow \mathrm{var}(X) = 0$

- $\mathrm{var}\left(c(X+a)\right) = c^2\ \mathrm{var}(X)$

Discrete Random Vectors

- The probability function of a discrete random vector: $(X_1, \ldots, X_n)$ is given by $P_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots X_n = x_n)$

Properties:

- $P_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \geq 0$
- $\sum_{X_1} \ldots \sum_{X_n} P_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = 1$
- $\sum_{X_{j \neq i}} P_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P_{X_i}(x_i)$

Continuous Random Vectors

- The joint probability distribution function of a continuous random vector: $(X_1, \ldots, X_n)$ is given by $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$.

Properties:

- $P(X_1, \ldots, X_n \in A) = \int \cdots \int_A f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \cdots dx_n$
- $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \geq 0$
- $\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n = 1$
- $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n = f_{X_i}(x_i)$

- Comment (continuous and discrete case): from the joint probability distribution function we can calculate the marginal probability distribution functions. But, the converse does not hold.

Cumulative Distribution Function of a Random Vector

- $F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P(X_1 \le x_1,\ldots,X_n \le x_n)$
- $F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) =$ $\int_{-\infty}^{x_1}\ldots\ \int_{-\infty}^{x_n} f_{X_1,\ldots,X_n}(u_1,\ldots,u_n)du_1\ldots du_n$
- $\frac{\partial^n F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)}{\partial_{x_1}\ldots\partial_{x_n}} = f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$

Properties:

- $\lim_{x(or\ \ y)\to-\infty} F_{x,y} = 0$
- $\lim_{y\to\infty} F_{x,y} = P(X < x) = F_X(x)$
- $\lim_{x\to\infty} F_{x,y} = P(Y < y) = F_Y(y)$

Independent Random Vector

- $F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdots F_{X_n}(x_n)$
- In the continuous case: $X_1 \cdots X_n$ are independent i.f.f. $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$
- In the discrete case: $X_1 \cdots X_n$ are independent i.f.f. $P_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} P_{X_i}(x_i)$

Properties of the Expected Value of Random Vector

- Let $X = (X_1, \ldots, X_n)$ be a random vector with density $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$, and consider the transformation $Y = h(x_1, \ldots, x_n)$, then:

- $\mathrm{E}Y = \int \ldots \int h(x_1, \ldots, x_n) f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n$

- $\mathrm{E}(X + Y) = \mathrm{E}X + \mathrm{E}Y$

- If X and Y are uncorrelated, then $\mathrm{E}(X \cdot Y) = \mathrm{E}X \cdot \mathrm{E}Y$

Transformation of a Random Vector

- Let $X = (X_1, \ldots, X_n)$ be a random vector with density $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$, consider the transformation $T(X_1, \ldots, X_n) = (Y_1, \ldots, Y_n)$, and the inverse transformation S, then:

  $$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = |J_s(y_1, \ldots, y_n)|\, f_{X_1,\ldots,X_n}(S(y_1, \ldots, y_n)),$$
  where J is the Jacobian off he transformation

- If $n = 2$, and we have the transformation $(X, Y) \Rightarrow TU, V)$, then:
  $$|J_S(U,V)| = \begin{vmatrix} \frac{\partial X}{\partial U} & \frac{\partial X}{\partial V} \\ \\ \frac{\partial Y}{\partial U} & \frac{\partial Y}{\partial V} \end{vmatrix} = |J_T(U,V)|^{-1} = \begin{vmatrix} \frac{\partial U}{\partial X} & \frac{\partial U}{\partial Y} \\ \\ \frac{\partial V}{\partial X} & \frac{\partial V}{\partial Y} \end{vmatrix}^{-1}$$

Conditional Distribution

- Let A be an event with $P(A) > 0$, and X be a random variable, then: $F_{X|A}(x|A) = P(X \leq x|A) = \frac{P(X \leq x \cap A)}{P(A)}$
- Definition: $f_{X|A}(x|A) = \frac{d}{dx} F_{X|A}(x|A)$
- $\mathrm{E}(x|A) = \int_{-\infty}^{\infty} x f_{X|A}(x|A)$
- If (X, Y) is a random vector, then $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
- $f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x) dx$
- Bayes thm: $f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y) f_Y(y)}{f_X(x)}$
- useful formula: $\mathrm{E}\left(\mathrm{E}(X|Y)\right) = \mathrm{E}X$

Covariance

- Definition: $\sigma_{X,Y} = \mathrm{cov}(X,Y) = \mathrm{E}[(X-\mu_X)(Y-\mu_Y)$, assuming that: $\mathrm{E}X^2 < \infty$ and $\mathrm{E}Y^2 < \infty$.

Properties:

- $\mathrm{cov}(X,X) = \mathrm{var}(X) > 0$
- $\mathrm{cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}X\mathrm{E}Y \Rightarrow \mathrm{if}\, X, Y\, \mathrm{indep.} \Rightarrow \mathrm{cov}(X,Y) = 0$
- $\mathrm{cov}(aX_1 + X_2 + c, Y) = a \cdot \mathrm{cov}(X_1, Y) + \mathrm{cov}(X_2, Y)$
- $\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$
- $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X,Y)$

Correlation

- Definition: $\mathrm{corr}(X, Y) = \rho_{X,Y} = \frac{\mathrm{cov} X,Y}{\sigma_X \sigma_Y}$, assuming $\sigma_X > 0, \sigma_Y > 0$

Properties:

- $-1 \leq \mathrm{corr}(X, Y) \leq 1$
- $\mathrm{corr}(X, Y) = \mathrm{corr}(Y, X)$
- $\mathrm{corr}(aX + c, Y) = \mathrm{sign}(a) \cdot \mathrm{corr}(X, Y)$

Multivariate Gaussian Vector

- Let $\mathbf{X} = [X_1, \ldots, X_p]^T$ be a random vector
- Definition: $\mathrm{E}(X_i) = \int \ldots \int x_i f_{X_1,\ldots,X_p}(x_1, \ldots, x_p) dx_1 \ldots dx_p$
- The expected value of $\mathbf{X}$ is given by:
  $\mu_{\mathbf{p}} = \mathrm{E}\mathbf{X} = [\mathrm{E}X_1, \ldots, \mathrm{E}X_p]^T$
- Definition: The covariance matrix $\Sigma_{p \times p}$ of $\mathbf{X}$ is defined as:

$$\begin{bmatrix} \mathrm{var}(X_1) & \mathrm{cov}(X_1, X_2) & \ldots & \mathrm{cov}(X_1, X_p) \\ \mathrm{cov}(X_2, X_1) & \mathrm{var}(X_2) & \ldots & \mathrm{cov}(X_2, X_p) \\ \vdots & \ddots & \ddots & \vdots \\ \mathrm{cov}(X_p, X_1) & \mathrm{cov}(X_p, X_2) & \ldots & \mathrm{var}(X_p) \end{bmatrix}$$

  Properties: $\Sigma_{p \times p}$ is symmetric and positive semidefinite (see linear algebra review).

Multivariate Gaussian Vector

- Definition 1: $\mathbf{X}$ is a multivariate Gaussian vector $\mathbf{X} \sim N_p(\mu_p, \Sigma_{p \times p})$ if the p.d.f. can be written as:

- $f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} exp\left(\frac{-1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right)$

- Definition 2: $\mathbf{X}$ is a multivariate Gaussian vector if for every $\mathbf{a} \in \mathbb{R}^p - 0$, $\mathbf{a}^T\mathbf{X}$ is a Gaussian random variable.

- We get: $\mathbf{a}^T\mathbf{X} \sim \mathcal{N}(\mathbf{a}^T\mu_p, \mathbf{a}^T\Sigma\mathbf{a})$

Multivariate Gaussian Vector

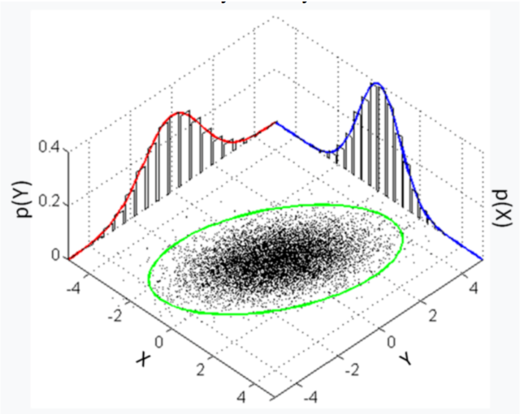


Figure 2: **Multivariate Gaussian vector** Diagram illustrating the joint and marginal probability distribution functions.

Identically Independently Distributed (i.i.d.) Random Variables & the Central Limit Theorem

- Consider $X_1, \ldots X_n$, such that every X_i has the same distribution and they are all independent.
- In particular: $\mathrm{E}X_i = \mu$ and var$X_i = \sigma^2$
- Why is the Gaussian distribution so popular? (Besides it's convenient mathematical formulation)
- Define: $S_n = \sum_{i=1}^{n} X_i$, $Y_n = \frac{S_n}{n}$
- Then (show at home): $\mathrm{E}Y_n = \mu$ and var$(Y_n) = \frac{\sigma^2}{n}$
- Central limit theorem:

$$\lim_{n \to \infty} \frac{S_n - n\mu}{\sqrt{n}} = \mathcal{N}(0, 1)$$

Interpretation: The average of several i.i.d. random variables has a normal distribution.

Central Limit Theorem

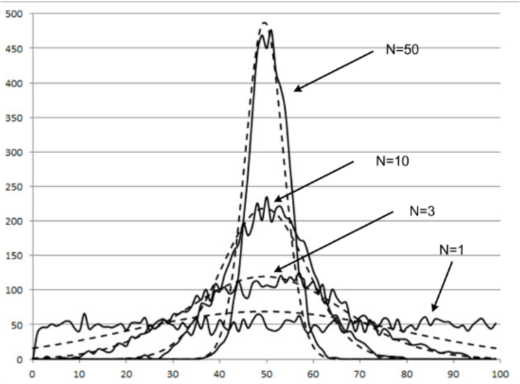


Figure 3: **Central limit theorem** The average of i.i.d samples has approaches Gaussian distribution as $n \to \infty$.