ECE276A: Sensing & Estimation in Robotics
Lecture 2: Probability Theory (Review)

Instructor:

Nikolay Atanasov: natanasov@ucsd.edu

Teaching Assistants:

Mo Shan: moshan@eng.ucsd.edu

Arash Asgharivaskasi: aasghari@eng.ucsd.edu

UC San Diego

JACOBS SCHOOL OF ENGINEERING

Electrical and Computer Engineering

Events

- **Experiment**: any procedure that can be repeated infinitely and has a well-defined set of possible outcomes.
- **Sample space** Ω: the set of possible outcomes of an experiment.
  - $\Omega = \{HH, HT, TH, TT\}$
  - $\Omega = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$
- **Event** A: a subset of the possible outcomes Ω
  - $A = \{HH\},\ B = \{HT, TH\}$
- **Probability of an event**: $\mathbb{P}(A) = \frac{\text{volume of } A}{\text{volume of all possible outcomes } \Omega}$

Measure and Probability Space

- σ-**algebra**: a collection of subsets of Ω closed under complementation and countable unions.
- **Borel σ-algebra** $\mathcal{B}$: the smallest σ-algebra containing all open sets from a topological space. Necessary because there is no valid translation invariant way to assign a finite measure to all subsets of $[0, 1)$.
- **Measurable space**: a tuple $(\Omega, \mathcal{F})$, where Ω is a sample space and $\mathcal{F}$ is a σ-algebra.
- **Measure**: a function $\mu : \mathcal{F} \to \mathbb{R}$ satisfying $\mu(A) \geq 0$ for all $A \in \mathcal{F}$ and countable additivity $\mu\left(\cup_i A_i\right) = \sum_i \mu(A_i)$ for disjoint A_i.
- **Probability measure**: a measure that satisfies $\mu(\Omega) = 1$.
- **Probability space**: a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a sample space, $\mathcal{F}$ is a σ-algebra, and $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a probability measure.

Probability Axioms

- **Probability Axioms**:
  - $\mathbb{P}(A) \geq 0$
  - $\mathbb{P}(\Omega) = 1$
  - If $\{A_i\}$ are disjoint, i.e., $A_i \cap A_j = \emptyset$, $\forall i \neq j$, then $\mathbb{P}(\bigcup_i A_i) = \sum_i \mathbb{P}(A_i)$
- **Corollary**:
  - $\mathbb{P}(\emptyset) = 0$
  - $\max\{\mathbb{P}(A), \mathbb{P}(B)\} \leq \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \leq \mathbb{P}(A) + \mathbb{P}(B)$
  - $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$

Events Example

- An experiment consists of randomly selecting one chip among ten chips marked $1, 2, 2, 3, 3, 3, 4, 4, 4, 4$.
  - What is a reasonable sample space for this experiment? $\Omega = \{1, 2, 3, 4\}$
  - What is the probability of observing a chip marked with an even number?

$$\mathbb{P}(\{2,4\}) = \mathbb{P}(\{2\} \cup \{4\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) = \frac{6}{10}$$

  - What is the probability of observing a chip marked with a prime number?

$$\mathbb{P}(\{2,3\}) = \mathbb{P}(\{2\} \cup \{3\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{3\}) = \frac{5}{10}$$

Set of Events

- **Conditional Probability**: $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$
- **Bayes Theorem**: assume $\mathbb{P}(B) > 0$

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

- **Total Probability**: If $\{A_1, \ldots, A_n\}$ is a partition of Ω, i.e., $\Omega = \bigcup_i A_i$ and $A_i \cap A_j = \emptyset$, $\forall i \neq j$, then:

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i)$$

- **Corollary**: If $\{A_1, \ldots, A_n\}$ is a partition of Ω, then:

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{n} \mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}$$

- **Independent events**: $\mathbb{P}\left(\bigcap_i A_i\right) = \prod_i \mathbb{P}(A_i)$
  - observing one does not give any information about another
  - in contrast, disjoint events never occur together: one occuring tells you that others will not occur and hence, <u>disjoint events are always dependent</u>

Independent Events Example

- A box contains 7 green and 3 red chips.
- Experiment: select one chip, replace the drawn chip, and repeat until the color red has been observed four times
- Assuming that no draw affects or is affected by any other draw, what is the probability that the experiment terminates on the ninth draw?

Independent Events Example

- Let the sample space Ω be a countably infinite set of all ordered tuples with elements from $\{r, g\}$:

$$\Omega = \{(r), (g), (r, r), (r, g), (g, r), (g, g), (r, r, r), \ldots\}$$

- Let $E \subset \Omega$ be such that:
  - Each tuple $e \in E$ has 9 components $e_1, \ldots, e_9$
  - The last component e_9 of each tuple $e \in E$ is r
  - There are exactly four components of r in each tuple $e \in E$

$$\text{Example: } (g, r, g, r, g, r, g, g, r) \in E$$

- Idea:
  - Show that every singleton subset $\{e\}$ of E has the same probability p_e
  - Determine the cardinality of E so that $\mathbb{P}(E) = \sum_{e \in E} \mathbb{P}(e) = |E| p_e$

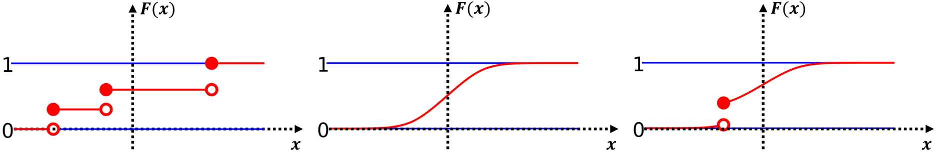
- Due to independence, for any element $e \in E$ we have:

$$\mathbb{P}(\{e\}) = \mathbb{P}\left(\{e_1\} \cap \{e_2\} \cap \cdots \cap \{e_9\}\right) = \prod_{i=1}^{9} \mathbb{P}(\{e_i\}) = \left(\frac{3}{10}\right)^4 \left(\frac{7}{10}\right)^5$$

- Since $e_9 = r$ for all $e \in E$, the cardinality of E is the number of ways to distribute 3 red chips among 8 slots, i.e., $|E| = \binom{8}{3}$

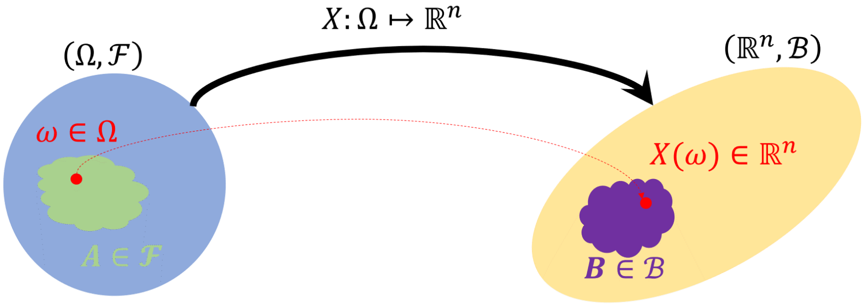
Random Variable

- **Random variable** X: an $\mathcal{F}$-measurable function from $(\Omega, \mathcal{F})$ to $(\mathbb{R}^n, \mathcal{B})$, i.e., a function $X : \Omega \to \mathbb{R}^n$ s.t. the preimage of every set in $\mathcal{B}$ is in $\mathcal{F}$.
- The **cumulative distribution function** (CDF) $F(x) := \mathbb{P}(X \leq x)$ of a random variable X is non-decreasing, right-continuous, and $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$.



(a) Discrete CDF (b) Continuous CDF (c) Mixed CDF

Random Variable



$\mathbb{P}: \mathcal{F} \mapsto \mathbb{R}$

$$\mathbb{P}(X \in B) = \mathbb{P}(A = \{\omega \in \Omega \mid X(\omega) \in B\})$$

"Volume of the preimage of B *under* X*"*

$$F_X(b) = \mathbb{P}(X \leq b) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in (-\infty, b_1] \times \cdots \times (-\infty, b_n]\})$$
$$= \int_{-\infty}^{b_n} \cdots \int_{-\infty}^{b_1} p_X(x_1, \dots, x_n) dx_1 \dots dx_n$$

CDF Examples

- $X \sim \mathcal{U}([a, b])$

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

- $X \sim \mathcal{U}(\{a, b\})$

$$F(x) = \begin{cases} 0 & x < a \\ 1/2 & a \leq x < b \\ 1 & x \geq b \end{cases}$$

- $X \sim Exp(\lambda)$ with $\lambda > 0$

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

- $X \sim \mathcal{N}(\mu, \sigma^2)$

$$F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}\right) dy$$

Probability Mass Function

- The **probability mass function** (pmf) $p(i)$ of a discrete random variable $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{Z}, 2^{\mathbb{Z}}, \mathbb{P} \circ X^{-1})$ satisfies:
  - $p(i) \geq 0$
  - $\sum_{i \in \mathbb{Z}} p(i) = 1$
  - $F(i) = \mathbb{P}(X \leq i) = \sum_{j \leq i} p(j)$
  - $\mathbb{P}(X = i) = p(i) \in [0, 1]$
  - $\mathbb{P}(a < X \leq b) = F(b) - F(a) = \sum_{a < j \leq b} p(j)$

Probability Density Function

- The **probability density function** (pdf) $p(x)$ of a continuous random variable $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}, \mathbb{P} \circ X^{-1})$ satisfies:
  - $p(x) \geq 0$
  - $\int p(y)dy = 1$
  - $F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} p(y)dy$
  - $\mathbb{P}(X = x) = \lim_{\epsilon \to 0} \int_{x}^{x+\epsilon} p(y)dy = 0$
  - $\mathbb{P}(a < X \leq b) = F(b) - F(a) = \int_{a}^{b} p(y)dy$
- Intuition:
  - The pdf $p(x)$ of X behaves like a derivative of the CDF $F(x)$
  - The values $p(a)$, $p(b)$ measure the relative likelihood of X being a or b
  - A discrete random variable $X \in \mathbb{Z}$ with pmf $m(i)$ can be viewed as continuous by defining its pdf as $p(x) := \sum_{i \in \mathbb{Z}} m(i)\delta(x - i)$, where δ is the Dirac delta function:
    $$\delta(x) := \begin{cases} \infty & x = 0 \\ 0 & x \neq 0 \end{cases} \qquad \int_{-\infty}^{\infty} \delta(x)dx = 1$$

pmf/pdf Examples

- $X \sim \mathcal{U}([a, b])$

$$p(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & x > b \end{cases}$$

- $X \sim \mathcal{U}(\{a, b\})$

$$p(i) = \begin{cases} \frac{1}{2} & i \in \{a, b\} \\ 0 & \text{else} \end{cases}$$

- $X \sim Exp(\lambda)$ with $\lambda > 0$

$$p(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases}$$

- $X \sim \mathcal{N}(\mu, \sigma^2)$

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

Expectation and Variance

- Given a random variable X with pdf p and a measurable function g, the **expectation** of $g(X)$ is:

$$\mathbb{E}\left[g(X)\right] = \int g(x)p(x)dx$$

- The **variance** of $g(X)$ is:

$$\begin{aligned} Var[g(X)] &= \mathbb{E}\left[\left(g(X) - \mathbb{E}[g(X)]\right)\left(g(X) - \mathbb{E}[g(X)]\right)^\top\right] \\ &= \mathbb{E}\left[g(X)g(X)^\top\right] - \mathbb{E}[g(X)]\mathbb{E}[g(X)]^\top \end{aligned}$$

- The **variance** of a sum of random variables is:

$$Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i] + \sum_{i=1}^{n}\sum_{j\neq i} Cov[X_i, X_j]$$

$$Cov[X_i, X_j] = \mathbb{E}\left[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])^\top\right] = \mathbb{E}\left[X_i X_j^\top\right] - \mathbb{E}[X_i]\mathbb{E}[X_j]^\top$$

Expectation and Variance Examples

- $X \sim \mathcal{U}([a, b])$

$$\mathbb{E}[X] = \int yp(y)dy = \frac{1}{b-a}\int_a^b ydy = \frac{b^2-a^2}{2(b-a)} = \frac{1}{2}(a+b)$$

$$Var[X] = \int y^2 p(y)dy - \mathbb{E}[X]^2 = \frac{b^3-a^3}{3(b-a)} - \frac{1}{4}(a+b)^2 = \frac{1}{12}(b-a)^2$$

- $X \sim \mathcal{U}(\{a, b\})$

$$\mathbb{E}[X] = \sum_{i\in\{a,b\}} i\ p(i) = \frac{1}{2}(a+b)$$

$$Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{2}(a^2+b^2) - \frac{1}{4}(a+b)^2 = \frac{1}{4}(b-a)^2$$

Expectation and Variance Examples

- $X \sim Exp(\lambda)$ with $\lambda > 0$

$$\mathbb{E}[X] = \int_0^\infty y\lambda e^{-\lambda y} dy \xlongequal{z=\lambda y,\, dz=\lambda dy} \frac{1}{\lambda}\int_0^\infty z e^{-z} dz$$

$$\xlongequal[du=dz,\, v=-e^{-z}]{u=z,\, dv=e^{-z}dz} \frac{1}{\lambda}\left(\left(-ze^{-z}\right)\Big|_0^\infty + \int_0^\infty e^{-z} dz\right) = \frac{1}{\lambda}(0+1) = \frac{1}{\lambda}$$

$$Var[X] = \int_0^\infty y^2 \lambda e^{-\lambda y} dy - \frac{1}{\lambda^2} \xlongequal{z=\lambda y,\, dz=\lambda dy} \frac{1}{\lambda^2}\left(\int_0^\infty z^2 e^{-z} dz - 1\right)$$

$$\xlongequal[du=2zdz,\, v=-e^{-z}]{u=z^2,\, dv=e^{-z}dz} \frac{1}{\lambda^2}\left(\left(-z^2e^{-z}\right)\Big|_0^\infty + 2\int_0^\infty e^{-z} dz - 1\right) = \frac{1}{\lambda^2}$$

- $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\mathbb{E}[X-\mu] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty \frac{(y-\mu)}{\sigma} \exp\left(-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}\right) dy$$

$$\xlongequal[dz=\frac{(y-\mu)}{\sigma}dy]{z=\frac{(y-\mu)^2}{2\sigma}} \frac{1}{\sqrt{2\pi}}\left(\int_\infty^{\mu^2/2\sigma} e^{-z/\sigma} dz + \int_{\mu^2/2\sigma}^\infty e^{-z/\sigma} dz\right) = 0$$

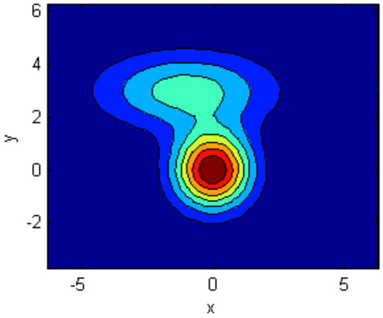
Set of Random Variables

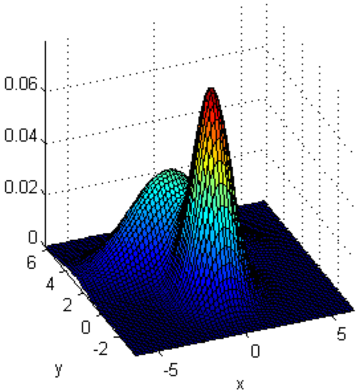
- The **joint distribution** of random variables $\{X_i\}_{i=1}^n$ on $(\Omega, \mathcal{F}, \mathbb{P})$ defines their simultaneous behavior and is associated with a cumulative distribution function $F(x_1, \ldots, x_n) := \mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n)$
- The CDF $F_i(x_i)$ of X_i defines its **marginal distribution**
- The random variables $\{X_i\}_{i=1}^n$ are **jointly independent** iff for all $\{A_i\}_{i=1}^n \subset \mathcal{F}$, $\mathbb{P}(\cap_{i=1}^n \{X_i \in A_i\}) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i)$
- Let X and Y be random variables and suppose $\mathbb{E}[X]$, $\mathbb{E}[Y]$, and $\mathbb{E}[XY]$ exist. Then, X and Y are **uncorrelated** iff $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ or equivalently $Cov[X, Y] = 0$.
- Independence implies uncorrelatedness

Gaussian Distribution

- **Gaussian random vector** $X \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$
  - parameters: **mean** $\boldsymbol{\mu} \in \mathbb{R}^n$, **covariance** $\Sigma \in \mathbb{S}^n_{\succ 0}$ (symmetric positive definite $n \times n$ matrix)
  - pdf: $\phi(\mathbf{x}; \boldsymbol{\mu}, \Sigma) := \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$
  - expectation: $\mathbb{E}[X] = \int \mathbf{x}\phi(\mathbf{x}; \boldsymbol{\mu}, \Sigma) d\mathbf{x} = \boldsymbol{\mu}$
  - variance: $Var[X] = \mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top\right] = \Sigma$
- **Gaussian mixture** $X \sim \mathcal{NM}(\{\alpha_k\}, \{\boldsymbol{\mu}_k\}, \{\Sigma_k\})$
  - parameters: **weights** $\alpha_k \geq 0$, $\sum_k \alpha_k = 1$,
    means $\boldsymbol{\mu}_k \in \mathbb{R}^n$, **covariances** $\Sigma_k \in \mathbb{S}^n_{\succeq 0}$
  - pdf: $p(\mathbf{x}) := \sum_k \alpha_k \phi(\mathbf{x}; \boldsymbol{\mu}_k, \Sigma_k)$
  - expectation: $\mathbb{E}[X] = \int \mathbf{x} p(\mathbf{x}) d\mathbf{x} = \sum_k \alpha_k \boldsymbol{\mu}_k =: \bar{\boldsymbol{\mu}}$
  - variance: $Var[X] = \mathbb{E}\left[XX^\top\right] - \mathbb{E}[X]\mathbb{E}[X]^\top = \sum_k \alpha_k \left(\Sigma_k + \boldsymbol{\mu}_k \boldsymbol{\mu}_k^\top\right) - \bar{\boldsymbol{\mu}}\bar{\boldsymbol{\mu}}^\top$

pdf of a Mixture of Two 2-D Gaussians





Example

- Suppose $V = (X, Y)$ is a continuous random vector with density $p_V(x, y) = 8xy$ for $0 < y < x$ and $0 < x < 1$
- Let $g(x, y) := 2x + y$
  - Determine $\mathbb{E}[g(V)]$
  - Evaluate $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ by finding the marginal densities of X and Y and then evaluating the appropriate univariate integrals
  - Determine $Var[g(V)]$

Example

$$\mathbb{E}\left[2X+Y\right] = \int_0^1 \int_0^x (2x+y)8xy\ dydx = \frac{32}{15}$$

$$p_X(x) = \int_0^x 8xy\ dy = 4x^3 \text{ for } 0 \le x \le 1$$

$$\mathbb{E}\left[X\right] = \int_0^1 xp_X(x)dx = \int_0^1 4x^4 dx = \frac{4}{5}$$

$$p_Y(y) = \int_y^1 8xy\ dx = 4y - 4y^3 \text{ for } 0 \le y \le 1$$

$$\mathbb{E}\left[Y\right] = \int_0^1 yp_Y(y)dy = \int_0^1 4y^2 - 4y^4 dy = \frac{8}{15}$$

$$Var\left[g(V)\right] = \mathbb{E}\left[(g(V) - \mathbb{E}\left[g(V)\right])^2\right] = \mathbb{E}\left[\left(2X + Y - \frac{32}{15}\right)^2\right]$$

$$= \int_0^1 \int_0^x \left(2x + y - \frac{32}{15}\right)^2 8xy\ dydx = \frac{17}{75}$$

Change of Density

- **Convolution**: Let X and Y be independent random variables with pdfs p and q, respectively. Then, the pdf of $Z = X + Y$ is given by the convolution of p and q:

$$[p * q](z) := \int p(z-y)q(y)dy = \int p(x)q(z-x)dx$$

- **Change of Density**: Let $Y = f(X)$. Then, with $dy = \left|\det\left(\frac{df}{dx}(x)\right)\right| dx$:

$$\begin{aligned}\mathbb{P}(Y \in A) = \mathbb{P}(X \in f^{-1}(A)) &= \int_{f^{-1}(A)} p_x(x)dx \\ &= \int_A \underbrace{\frac{1}{\left|\det\left(\frac{df}{dx}(f^{-1}(y))\right)\right|} p_x(f^{-1}(y))}_{p_y(y)} dy\end{aligned}$$

Change of Density Example

- Let $X \sim \mathcal{N}(0, \sigma^2)$ and $Y = f(X) = \exp(X)$
- Note that $f(x)$ is invertible $f^{-1}(y) = \log(y)$
- The infinitesimal integration volumes for y and x are related by:

$$dy = \left| \det \left(\frac{df}{dx}(x) \right) \right| dx = \exp(x) dx$$

- Using change of density with $A = [0, \infty)$ and $f^{-1}(A) = (-\infty, \infty)$:

$$\mathbb{P}(Y \in [0, \infty)) = \int_{-\infty}^{\infty} \phi(x; 0, \sigma^2) dx = \int_0^{\infty} \frac{1}{\exp(\log(y))} \phi(\log(y); 0, \sigma^2) dy$$
$$= \int_0^{\infty} \underbrace{\frac{1}{y} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{\log^2(y)}{\sigma^2} \right)}_{p(y)} dy$$

Change of Density Example

- Let $V := (X, Y)$ be a random vector with pdf:

$$p_V(x, y) := \begin{cases} 2y - x & x < y < 2x \text{ and } 1 < x < 2 \\ 0 & \text{else} \end{cases}$$

- Let $T := (M, N) = g(V) := \left(\frac{2X-Y}{3}, \frac{X+Y}{3}\right)$ be a function of V
- Note that $X = M + N$ and $Y = 2N - M$ and, hence, the pdf of V is non-zero for $0 < m < n/2$ and $1 < m + n < 2$. Also:

$$\det\left(\frac{dg}{dv}\right) = \det\begin{bmatrix} 2/3 & -1/3 \\ 1/3 & 1/3 \end{bmatrix} = \frac{1}{3}$$

- The pdf T is:

$$p_T(m, n) = \begin{cases} \frac{1}{\left|\det\left(\frac{dg}{dv}(m+n, 2n-m)\right)\right|} p_V(m + n, 2n - m), & \begin{array}{l} 0 < m < n/2 \text{ and} \\ 1 < m + n < 2, \end{array} \\ 0, & \text{else.} \end{cases}$$

Conditional and Total Probability

- **Total Probability**: If two random variables X, Y have a joint pdf $p(x, y)$, the marginal pdf $p(x)$ of X is:

$$p(x) = \int p(x, y) dy$$

- **Conditional Distribution**: If two random variables X, Y have a joint pdf $p(x, y)$, the pdf $p(x|y)$ of X conditioned on $Y = y$ and the pdf $p(y|x)$ of Y conditioned on $X = x$ satisfy

$$p(x, y) = p(x|y)p(y) = p(y|x)p(x)$$

- **Bayes Theorem**: The pdf $p(x|y)$ of X conditioned on $Y = y$ can be expressed in terms of the pdf $p(y|x)$ of Y conditioned on $X = x$ and the marginal pdf $p(x)$ of X:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\int p(y \mid x')p(x')dx'}$$

Conditional Probability Example

- Suppose that $V = (X, Y)$ is a discrete random vector with probability mass function:

$$p_V(x,y) = \begin{cases} 0.10 & \text{if } (x,y) = (0,0) \\ 0.20 & \text{if } (x,y) = (0,1) \\ 0.30 & \text{if } (x,y) = (1,0) \\ 0.15 & \text{if } (x,y) = (1,1) \\ 0.25 & \text{if } (x,y) = (2,2) \\ 0 & \text{elsewhere} \end{cases}$$

  - What is the conditional probability that V is $(0,0)$ given that V is $(0,0)$ or $(1,1)$?
  - What is the conditional probability that X is 1 or 2 given that Y is 0 or 1?
  - What is the probability that X is 1 or 2?
  - What is the probability mass function of $X \mid Y = 0$?
  - What is the expected value of $X \mid Y = 0$?

Conditional Probability Example

$$\mathbb{P}\left(V \in \{(0,0)\} \mid V \in \{(0,0),(1,1)\}\right) = \frac{\mathbb{P}\left(V \in \{(0,0)\} \cap \{(0,0),(1,1)\}\right)}{\mathbb{P}\left(V \in \{(0,0),(1,1)\}\right)}$$

$$= \frac{0.10}{0.25} = 0.4$$

$$\mathbb{P}\left(X \in \{1,2\} \mid Y \in \{0,1\}\right) = \mathbb{P}\left(V \in \{1,2\} \times \mathbb{R} \mid V \in \mathbb{R} \times \{0,1\}\right)$$

$$= \frac{\mathbb{P}\left(V \in \{(1,0),(1,1)\}\right)}{\mathbb{P}\left(V \in \{(0,0),(0,1),(1,0),(1,1)\}\right)} = \frac{0.45}{0.75} = 0.6$$

$$\mathbb{P}\left(X \in \{1,2\}\right) = \mathbb{P}\left(V \in \{1,2\} \times \mathbb{R}\right) = 0.7$$

$$p_{X|Y=0}(x) = \frac{p_V(x,0)}{\sum_{x' \in \{0,1\}} p_V(x',0)} = \frac{1}{0.4} p_V(x,0) = \begin{cases} 0.25 & \text{if } x = 0 \\ 0.75 & \text{if } x = 1 \end{cases}$$

$$\mathbb{E}\left[X \mid Y = 0\right] = \sum_{x \in \{0,1\}} x p_{X|Y=0}(x) = p_{X|Y=0}(1) = 0.75$$