Lecture 11

- **Read:** Chapter 8.1-8.2, 9.1.

Statistical Inference

- Significance Testing
- Binary Hypothesis Testing
  - Maximum A posteriori Probability (MAP) Test
  - Maximum Likelihood (ML) Test
- Estimation of a Random Variable
  - Blind Estimation of X
  - Estimation of X Given an Event
  - Estimation of X Given a Random Variable

Statistical Inference

- Need to be able to reason in the presence of uncertainty
- Analyze observations of an experiment to reach conclusions with some assessment of the quality or risk associated with these conclusions
- When the conclusion is based on the properties of random variables, the reasoning is referred to as a **statistical inference**.
- We will look at five categories of statistical inference:
  - Significance testing
  - Hypothesis testing
  - Estimation of a random variable
  - Point estimation of a model parameter
  - Confidence interval estimation of a model parameter

Statistical Inference

- Like probability theory, the theory of statistical inference refers to an experiment consisting of a procedure and observations.
- In each statistical inference method, there is a set of possible conclusions and a means of measuring the accuracy of a conclusion.
- A statistical inference method assigns a conclusion to each possible outcome of the experiment and consists of three steps:
  - Perform an experiment
  - Observe an outcome
  - State a conclusion
- The assignment of conclusions to outcomes is based on probability theory.
- The aim of the assignment is to achieve the highest possible accuracy.

Statistical Inference: Various Scenarios

- **Significance/Hypothesis Testing:**
  - **Significance:** Given a single hypothesis H_0, figure out if it holds.
  - **Hypothesis:** Given several hypotheses $H_1, H_2, ..., H_n$, which one is "best"?

- **Estimation:**
  - **Random variable:** Goal is to estimate an RV X, based on some observation (e.g., an event or another random variable).
  - **Parameter estimation:** Here, X, might be modeled using some parametric distribution (e.g., $X \sim exp(\lambda)$), and based on observation we wish to estimate the true parameter λ.

    1. point estimate $\rightarrow \hat{\lambda}$
    2. confidence interval estimate
       $\rightarrow [a, b]$
       where, $P[\lambda \in [a, b]] \geq \underbrace{1 - \alpha}_{\text{confidence}}$

Statistical Inference: Example

- $X_1, \ldots, X_n$ are iid samples of an exponential RV X.
- $E[X] = \lambda$ is unknown.
- Can we use $X_1, \ldots, X_n$ to learn about λ?

Statistical Inference: Significance Testing Example

- **Conclusion:** Accept or reject the hypothesis that the observations result from a probability model H_0.
- **Accuracy Measure:** Probability of rejecting the hypothesis when it is true
- **Question:** Assuming λ is a constant, should we accept or reject the hypothesis that $\lambda = 3.5$?

Statistical Inference: Hypothesis Testing Example

- **Conclusion:** The observations result from one of two hypothetical probability models (H_0 and H_1).
- **Accuracy Measure:** Probability that the conclusion is H_0, when the true model is H_1.
- **Question:** Assuming λ is a constant, which one does λ equal - 2.5 or 3.5?

Statistical Inference: Random Variable Estimation Example

- **Conclusion:** The value of random variable X is $\hat{X}$.
- **Accuracy Measure:** The mean square error: $E[(X - \hat{X})^2]$
- **Method:** Assume Λ is an RV with $f_\Lambda(\lambda)$ and the experiment that generates $X_1, ..., X_n$ implicitly has two parts:
    1. Generate a sample value $\Lambda = \lambda$.
    2. Generate each X_i from an exponential distribution with expected value λ.
- **Question:** What is $\hat{\lambda}$, the best estimate of λ?

Statistical Inference: Point Estimation of a Parameter

- **Conclusion:** The value of a parameter of a probability model (e.g., expected value) is $\hat{c}$.
- **Accuracy Measure:** The mean square error: $E[(c - \hat{c})^2]$ (where, c is the true value of the parameter.)
- **Question:** Assuming λ is a constant, what is $\hat{\lambda}$, the best estimate of λ?

Statistical Inference: Confidence Interval Estimation Example

- **Conclusion:** The value, c, of a parameter of an RV is in the interval $a \leq c \leq b$.
- **Accuracy Measure:** The interval size, $b - a$, and α, the probability that the conclusion is false.
- **Question:** Assuming λ is a constant, what values of λ_1 and λ_2 satisfy $P[\lambda_1 \leq \lambda \leq \lambda_2] \geq 0.95$?

Significance Testing: Setup

- We have a single hypothesis (probability model), H_0.
- We make some observation $X = x$.
- The test is designed so that it divides the set of possible observations S_X into two sets, A and R.

$$\text{If } x \in A \;\Rightarrow\; \text{We accept hypothesis } H_0.$$
$$\text{If } x \in R \;\Rightarrow\; \text{We reject hypothesis } H_0.$$

- Accuracy measure is P[false rejection].
- Let $\alpha = P[x \in R | H_0]$. That is, α is the probability that we reject the hypothesis when in fact it is correct.
- We will call α the "significance level" of the test.

Significance Testing: Two Types of Errors

- **Type 1: (false rejection)** reject H_0 when it is in fact true.
- **Type 2: (false acceptance)** accept H_0 when it is false.

$*$ With only one hypothesis, it is only possible to calculate the probability of a Type 1 error. We cannot measure the probability of false acceptance.

Binary Hypothesis Testing: Example

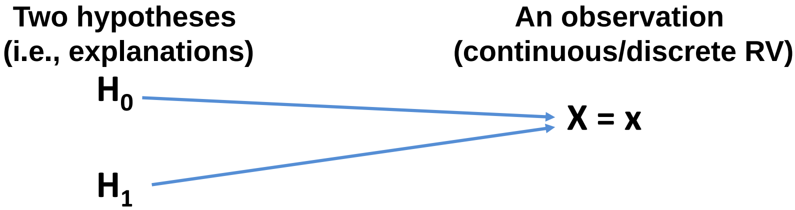
- **Observation:** $Y \in \{0, 1\}$
- **Hypothesis:** $H_0 = \text{“}\{X = 0\}\text{”}$
  $H_1 = \text{“}\{X = 1\}\text{”}$
- Given that $Y = 0$, which hypothesis (H_0 or H_1) should we select?

  $P[H_0|Y = 0]$
  $P[H_1|Y = 0]$

Maximum Likelihood (ML)
Pick biggest.

Maximum A Posteriori (MAP)
Pick biggest.
$P[Y = 0|H_0]$
$P[Y = 0|H_1]$

Binary Hypothesis Testing: Setup



Binary Hypothesis Testing: Model

- **Prior Probabilities:** $P[H_0]$, $P[H_1]$
  - These reflect the state of knowledge about the probability model before an outcome is observed.
  - These may or may not be known.

- **Likelihoods or Transition Probabilities:** likelihood of X given H_i
  - $P[X = x|H_i]$, $i = 0, 1$ [when the observation leads to a discrete RV, X]
  - $f_{X|H_i}(x)$, $i = 0, 1$ [when the observation leads to a continuous RV, X]

- **Goal:** For each observation $X = x$, we need to decide on a hypothesis $\hat{h}(x)$ such that

$$\hat{h}(x) = 0 \text{ or } 1$$

Binary Hypothesis Testing: Two Types of Errors

- **Type 1:** $P[\hat{h}(X) = 1|H_0]$: accepting H_1 when H_0 is true
- **Type 2:** $P[\hat{h}(X) = 0|H_1]$: accepting H_0 when H_1 is true
- **Note:** The "cost" of these two types of errors may differ.
  - **Example:** Missile detection:

    $H_0 =$ no missile
    $H_1 =$ presence of missile

    Type 1 = false alarm
    Type 2 = miss
- What criteria should be used to design hypothesis testing rules?

Binary Hypothesis Testing: Radar Test

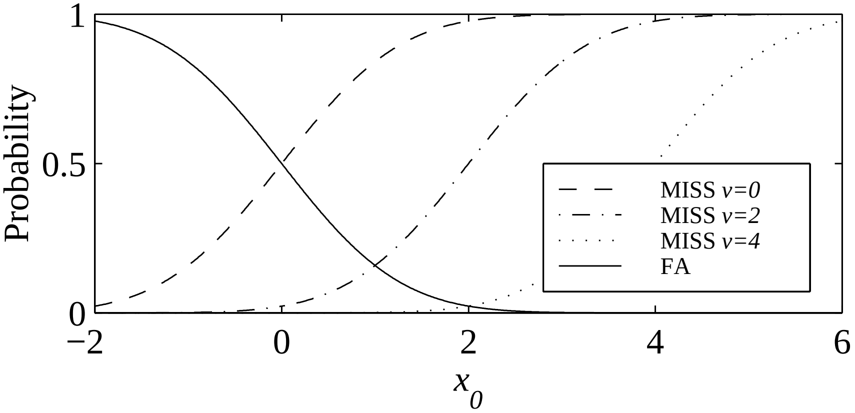
- Noise voltage in a radar detector is Gaussian RV N with $E[N] = 0$ volts and $Var[N] = \sigma^2$ $volts^2$.
- If target (H_1): Output is $X = \nu + N$
- No target (H_0): Output is $X = N$
- Periodically, the detector performs a binary hypothesis test
- Acceptance regions:

$$A_0 = \{X \leq x_0\},\ A_1 = \{X > x_0\}$$

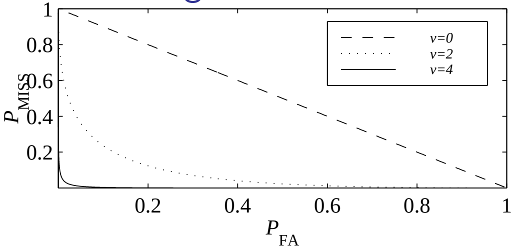
Binary Hypothesis Testing: Radar Test Errors (I)

$$
\begin{aligned}
P_{\text{miss}} &= P[A_0|H_1] \\
&= P[X \leq x_0|H_1] = \Phi\left(\frac{x_0 - \nu}{\sigma}\right) \\
P_{\text{false acceptance}} &= P[A_1|H_0] \\
&= P[X > x_0|H_0] = 1 - \Phi\left(\frac{x_0}{\sigma}\right)
\end{aligned}
$$

Binary Hypothesis Testing: Radar Test Errors (II)



Binary Hypothesis Testing: Radar Test Errors (III)



- When $\nu = 0$, the received signal is the same, regardless of whether or not a target is present.
- In this case, $P_{MISS} = 1 - P_{falsealarm}$.
- As ν increases, it is easier for the detector to distinguish between the two targets.
- We see that the curve for $\nu = 4$ is better than the curve for $\nu = 2$, which is better than the curve for $\nu = 0$.
- Therefore, we can choose a value of x_0 such that both P_{MISS} and $P_{falsealarm}$ are lower for $\nu = 4$ than for $\nu = 2$.

Binary Hypothesis Testing: Radar Test Probability of Error

- Conditional error probabilities

$$P_{\text{false acceptance}} = P[A_1|H_0] \quad , \quad P_{\text{miss}} = P[A_0|H_1]$$

- Total probability of error

$$P_{\text{ERROR}} = P[A_1|H_0]P[H_0] + P[A_0|H_1]P[H_1]$$

MAP Binary Hypothesis Test

- Binary Hypothesis Test: Observation s
- If we choose A_i,

$$P[\text{correct decision}|s] = P[H_i|s]$$

- Minimum P_{ERROR} rule
  - $s \in A_0$ if $P[H_0|s] \geq P[H_1|s]$
  - $s \in A_1$ if $P[H_1|s] > P[H_0|s]$
- **A posteriori probabilities:** $\mathrm{P}[H_i|s]$
  - Just as the a priori probabilities $P[H_0]$ and $P[H_1]$ reflect our knowledge of H_0 and H_1 prior to performing an experiment, $P[H_0|s]$ and $P[H_1|s]$ reflect our knowledge after observing s.

Likelihood Ratio Test

- By Bayes' Theorem,
$$P[H_i|s] = \frac{P[s|H_i]P[H_i]}{P[s]}$$
- The MAP Rule becomes
  - $s \in A_0$ if $\frac{P[s|H_0]P[H_0]}{P[s]} \geq \frac{P[s|H_1]P[H_1]}{P[s]}$
  - $s \in A_1$ if $\frac{P[s|H_1]P[H_1]}{P[s]} > \frac{P[s|H_0]P[H_0]}{P[s]}$

Likelihood Ratio with RV X

- When the sample space of the experiment is the range of the random variable X, we can express the MAP rule in terms of the conditional PMFs or PDFs of X.

- For an experiment that produces a continuous random variable X, the MAP binary hypothesis test is:
  - $x \in A_0$ if $\frac{f_{X|H_0}(x)}{f_{X|H_1}(x)} \geq \frac{P[H_1]}{P[H_0]}$
  - $x \in A_1$ if $\frac{f_{X|H_0}(x)}{f_{X|H_1}(x)} < \frac{P[H_1]}{P[H_0]}$

Maximum A Posteriori Criterion (MAP): Discrete X

- Choose hypothesis that maximizes
$$P[H_i|X = x] = \frac{P[X = x|H_i]P[H_i]}{P[X = x]}$$
i.e., maximize probability of hypothesis given observation
- $P[H_i|X = x]$, is the a posteriori probability of H_i. It reflects our knowledge after observing x.

$$\hat{h}_{MAP}(x) = \begin{cases} 0(H_0) & , P[X = x|H_0]P[H_0] \geq P[X = x|H_1]P[H_1] \\ 1(H_1) & , \text{otherwise} \end{cases}$$

$$\underbrace{\frac{P[X = x|H_0]}{P[X = x|H_1]}}_{\text{likelihood ratio}} \geq \underbrace{\frac{P[H_1]}{P[H_0]}}_{\text{threshold}}$$

- Likelihood ratio: Evidence, based on observation, in favor of H_0. If $> 1 \Rightarrow H_0$ more likely than H_1
- Threshold: Evidence, prior to performing the experiment, in favor of H_1

Maximum A Posteriori Criterion (MAP): Discrete X (cont.)

- **Theorem:** MAP criterion minimizes overall probability of error.

$$P_{error} = P[\hat{h}(X) = 1|H_0]P[H_0] + P[\hat{h}(X) = 0|H_1]P[H_1]$$

Maximum A Posteriori Criterion (MAP): Continuous X

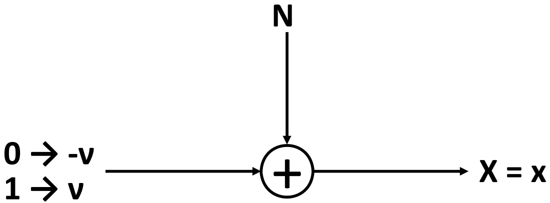
$$\hat{h}_{MAP}(x) = \begin{cases} 0 \quad (H_0) & \frac{f_{X|H_0}(x)}{f_{X|H_1}(x)} \geq \frac{P[H_1]}{P[H_0]} \\ 1 \quad (H_1) & \text{, otherwise} \end{cases}$$

MAP Example: Binary Communications System (I)

- With probability p, a binary communications system transmits a "0".
- It transmits a "1" with probability $1 - p$.
- The voltage $\nu \geq 0$ is the information component of the received signal and N is the noise component, where $N \sim N(0, \sigma^2)$.
- The received signal X is:

$$P[H_0] = p, \qquad \text{If } H_0\text{, then } X = -\nu + N$$

$$P[H_1] = 1 - p, \qquad \text{If } H_1\text{, then } X = \nu + N$$



MAP Example: Binary Communications System (II)

- Conditional probability density functions:

$$f_{X|H_0}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(x+\nu)^2}{2\sigma^2}\right]$$
$$f_{X|H_1}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(x-\nu)^2}{2\sigma^2}\right]$$

- Likelihood ratio: $x \in A_0$ if $\frac{f_{X|H_0}(x)}{f_{X|H_1}(x)} \geq \frac{P[H_1]}{P[H_0]}$

$$\frac{f_{X|H_0}(x)}{f_{X|H_1}(x)} = exp\left[-\frac{(x+\nu)^2}{2\sigma^2} + \frac{(x-\nu)^2}{2\sigma^2}\right] \overset{?}{\geq} \frac{1-p}{p}$$

MAP Example: Binary Communications System (III)

- Taking ln of both sides,

$$
\begin{aligned}
-\frac{(x+\nu)^2}{2\sigma^2} + \frac{(x-\nu)^2}{2\sigma^2} &\overset{?}{\geq} \ln\left(\frac{1-p}{p}\right) \\
-\frac{4x\nu}{2\sigma^2} &\overset{?}{\geq} \ln\left(\frac{1-p}{p}\right) \\
-\frac{2x\nu}{\sigma^2} &\overset{?}{\geq} \ln\left(\frac{1-p}{p}\right) \\
x &\overset{?}{\leq} \frac{\sigma^2}{2\nu} \ln\left(\frac{p}{1-p}\right)
\end{aligned}
$$

MAP Example: Binary Communications System (IV)

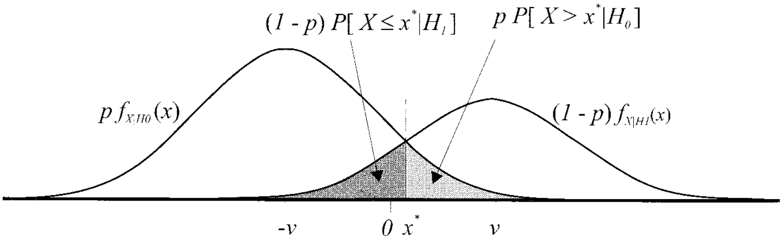
- So,

$$\hat{h}_{MAP}(x) = \begin{cases} 0(H_0) & , x \leq \overbrace{-\frac{\sigma^2}{2\nu} \ln\left(\frac{p}{1-p}\right)}^{x^*} \\ 1(H_1) & , \text{otherwise} \end{cases}$$

MAP Example: Binary Communications System (V)

- Error probabilities

$$P_{ERROR} = pP[X > x^*|H_0] + (1-p)P[X \leq x^*|H_1]$$



- The threshold x^* is the value of x for which the two likelihood functions, each multiplied by a prior probability, are equal.
- The probability of error equals the sum of the shaded areas.
- Compared to all other decision rules, the threshold x^* provides the minimum possible error.

MAP to ML

- Given an observation, we know that the MAP rule minimizes the probability that we accept the wrong hypothesis.
- However, the MAP rule requires that we know the a priori probabilities $P[H_i]$ of the competing hypotheses.
- In many situations, though, it is difficult to specify a priori probabilities.
- Instead, one might want to treat the hypothesis as some sort of "unknown" and choose a hypothesis H_i for which $P[s|H_i]$, the conditional probability of the outcome s given the hypothesis H_i is largest.

Maximum Likelihood (ML) Criterion

- Choose hypothesis that maximizes likelihood (to avoid making assumptions about the a priori probabilities $P[H_i]$).
- That is, choose the hypothesis that maximizes:

$$P[X = x|H_i],\ i = 0, 1$$
$$f_{X|H_i}(x),\ i = 0, 1$$

- That is, choose hypothesis most likely to give the current observation.

$$\hat{x}_{ML}(x) = \begin{cases} 0(H_0) & , P[X = x|H_0] \geq P[X = x|H_1] \\ & \quad (f_{X|H_0}(x) \geq f_{X|H_1}(x)) \\ 1(H_1) & , \text{otherwise} \end{cases}$$

$$\hat{x}_{ML}(x) = \begin{cases} 0(H_0) & , \frac{f_{X|H_0}(x)}{f_{X|H_1}(x)} \geq 1 \\ 1(H_1) & , \text{otherwise} \end{cases}$$

Maximum Likelihood (ML) Criterion: Remarks

- ML $=$ MAP if $P[H_0] = P[H_1]$
  - i.e., the ML test is the MAP test with $P[H_0] = P[H_1]$.
- If the prior information, i.e., $P[H_0]$ and $P[H_1]$ are not known, then we can use ML.
- If $P[H_0] = P[H_1]$, the ML minimizes the probability of error.
- In essence, by trying to avoid making a priori assumptions, the ML rule assumes that all hypotheses are equally likely.

Estimation: Applications

- Measurement, approximation, model fitting
- Pattern matching, learning theory, tracking, control, etc.

Estimation of a Random Variable

- Minimize $E[(X - \hat{x}_B)^2]$.

  $\hat{x}_B$

$$\begin{aligned} \frac{\partial}{\partial \hat{x}_B}(E[X^2] - 2E[X]\hat{x}_B + \hat{x}_B^2) &= 0 \\ -2E[X] + 2\hat{x}_B &= 0 \\ \hat{x}_B &= E[X] \end{aligned}$$

- Minimize $E[(X - \hat{X}_M(Y))^2]$.

  $\hat{x}_M(\cdot)$

Estimation of a Random Variable: Setup

- Experiment produces an RV X.
- However, we may be unable to observe X directly.
- Instead, we observe an event or RV that gives partial information about the sample value of X.
- Example: Noisy observations: $Y = X + N$
  - We would like to know X, but it has been corrupted by an RV, N, and we only have access to RV $Y = X + N$.
  - This problem arises all the time in communications systems.
- **Goal:** Find the "best" estimate $\hat{x}$ for X.
  $\Rightarrow$ Here, "best" will mean minimum expected error

$$\min \underbrace{E\left[(X - \hat{x})^2\right]}_{\text{mean square error}}$$

Estimation of a Random Variable: Estimation Accuracy

- Accuracy measure: mean square error (MSE)

$$e = E[(X - \hat{x})^2]$$

- MSE$=$ one of many definitions of accuracy
- Alternatives: $E[|X - \hat{x}|]$ or $\max\{|X - \hat{x}|\}$
- We focus on MSE which is the most widely used accuracy measure because it lends itself to mathematical analysis and often leads to estimates which are convenient to compute.

Estimation of a Random Variable: Types of Partial Information

- Case 1: Blind Estimation: probability model (e.g., PDF or PMF) of X
- Case 2: Given an Event: probability model of X and information that $x \in A$
- Case 3: Given a Random Variable: probability models of X and Y and information that $Y = y$

Estimation of a Random Variable: Case 1, Blind Estimation

- **Theorem:** Without observations, the minimum mean square error (MSE) estimate of X is $\hat{x}_B$:

$$\min_{\hat{x}} E[(X-\hat{x})^2] \overset{\frac{\partial}{\partial \hat{x}}()=0}{\Rightarrow} \hat{x}_B = E[X]$$

(B = blind estimation)

- **Alternative Proof:**

$$\begin{aligned} E[(X-\hat{x})^2] &= E[((X-\hat{x}_B)+(\hat{x}_B-\hat{x}))^2] \\ &= E[(X-\hat{x}_B)^2 + 2(X-\hat{x}_B)(\hat{x}_B-\hat{x}) + (\hat{x}_B-\hat{x})^2] \\ &= E[(X-\hat{x}_B)^2] + \underbrace{(\hat{x}_B-\hat{x})^2}_{\geq 0} \\ &\geq E[(X-\hat{x}_B)^2] \end{aligned}$$

Estimation of a Random Variable: Case 2, Given an Event

- Observation: $x \in A$
- Given A, X has a conditional PDF $f_{X|A}(x)$.
- Goal: Minimize $E[(X - \hat{x})^2|A]$.
- Same as blind estimation, but $f_{X|A}(x|A)$ (PDF of X given A occurred) replaces $f_X(x)$.
- **Theorem:** Given that an event $x \in A$ occurred, the best (minimum MSE) estimate $\hat{x}_A$ for X is

$$\min_{\hat{x}} E[(X - \hat{x})^2|A] \Rightarrow \hat{x}_A = E[X|A]$$

Estimation of a Random Variable Case 2: Example

- The duration T minutes of a phone call is an exponential RV with expected value, $E[T] = 3$ minutes.
- If we observe that a call has already lasted 2 minutes, what is the minimum mean square error estimate of the call duration?

. . . .

- $T \sim \exp(1/3)$ or

$$f_T(t) = \begin{cases} \frac{1}{3}e^{-1/3t} & , t \geq 0 \\ 0 & , \text{otherwise} \end{cases}$$

- If the call is still in progress after 2 minutes, we have $t \in A = \{T > 2\}$.
- Therefore, the minimum mean square error estimate of T is

$$\hat{t} = E[T|T > 2]$$

$$\begin{aligned} P[T > 2] = \int_2^{+\infty} f_T(t)dt &= 1 - P[T \leq 2] \\ &= 1 - (1 - e^{-2/3}) \\ &= e^{-2/3} \end{aligned}$$

Estimation of a Random Variable Case 2: Example

- The duration T minutes of a phone call is an exponential RV with expected value, $E[T] = 3$ minutes.
- If we observe that a call has already lasted 2 minutes, what is the minimum mean square error estimate of the call duration?

....

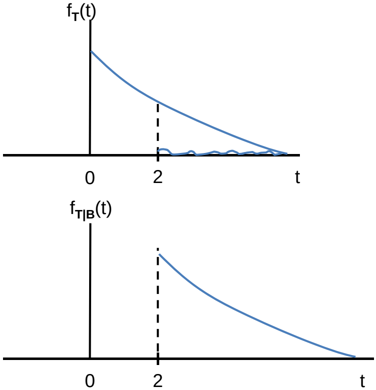
$$f_{T|B}(t) = \begin{cases} \frac{f_T(t)}{P[B]} = \frac{\frac{1}{3}e^{-1/3t}}{e^{-2/3}} = \frac{1}{3}e^{-\frac{1}{3}(t-2)} & , \ t > 2 \\ 0 & , \text{otherwise} \end{cases}$$

Estimation of a Random Variable: Case 2, Given an Event: Example (cont.)

- Find $E[T|T > 2]$.

. . . .

$$E[T|B] = \int_{-\infty}^{+\infty} t f_{T|B}(t) dt$$
$$= \int_{2}^{+\infty} t \frac{1}{3} e^{-\frac{1}{3}(t-2)} dt$$
$$= 5 \text{ minutes}$$



Reminder on integration by parts:

$$\int_a^b f(x)g'(x)dx = f(x)g(x)|_a^b - \int_a^b f'(x)g(x)dx$$

Estimation of a Random Variable: Case 3, Given a Random Variable (I)

- For observation $Y = y$, assign $\hat{x}$ near X.
- $\hat{x}$ is a function of sample value y: $\hat{x}_M(y)$
- $\hat{x}_M(y)$ is a sample value of the RV $\hat{X}_M(Y)$.
- The squared error is $(X - \hat{X}_M(Y))^2$, a function of RVs X and Y.

Estimation of a Random Variable: Case 3, Given a Random Variable (II)

- Given a random variable Y, find the best estimate for X.
- **Theorem:** Given $Y = y$, the best (minimum MSE) estimate for X is $\hat{x}_M(y)$:

$$\min_{\hat{x}} E[(X - \hat{x})^2 | Y = y] \Rightarrow \hat{x}_M(y) = E[X|Y = y]$$

Estimation of a Random Variable: Case 3, Given a Random Variable (III)

- **Theorem:** Suppose we would like to find the best estimator function of X given Y.
$$\min_{g(\cdot)} E[(X-g(Y))^2]$$

- Recall from the "Multiple Continuous Random Variables" lecture,
$$E[E[g(X,Y)|Y]] = \int_{-\infty}^{+\infty} E[g(X,Y)|Y=y]f_Y(y)dy = E[g(X,Y)]$$

- Note that:
$$E[(X-g(Y))^2] = \int_{-\infty}^{+\infty} E[(X-g(Y))^2|Y=y]f_Y(y)dy$$
(We can minimize these terms by letting $\underline{g(y)} = E[X|Y=y]$.)

- The optimal estimator function is given by $\hat{x}_M(\cdot)$.

Estimation of a Random Variable: Case 3, Given a Random Variable (IV)

- **Notation:** The best estimate for X given Y is

$$\hat{x}_M(Y) = E[X|Y] = \hat{X}_M(Y)$$

- **Theorem:** The conditional expectation of X given Y, $E[X|Y]$ is the minimum mean square error (MMSE) estimator for X given Y.

- **Example:**
  - R uniform on $[0, 1]$. Given $R = r$, X uniform on $[0, r]$.
  - Find $\hat{X}_M(R)$.

. . . .

  - From the problem statement,

$$f_{X|R}(x|r) = \begin{cases} \frac{1}{r} & , 0 \leq x \leq r \\ 0 & , \text{otherwise} \end{cases}$$

Estimation of a Random Variable: Case 3, Given a Random Variable (V)

- **Example:** continued:
  - The MMSE estimate of X, given $R = r$ is
  $$\hat{x}_M(R) = E[X|R = r] = \int_0^r x \cdot \frac{1}{r} dx = \frac{1}{r} \frac{x^2}{2} \Big|_{x=0}^{x=r} = r/2$$
  - The MMSE estimator is $\hat{X}_M(R) = R/2$.
  - If we want to find $\hat{R}_M(X)$, we need to find the conditional PDF $f_{R|X}(r|x)$:
  $$f_{R|X}(r|x) = \begin{cases} \frac{1}{-r \ln x} & , 0 \leq x \leq r \leq 1 \\ 0 & , \text{otherwise} \end{cases}$$
  - The MMSE estimator is $\hat{r}_M(X) = E[R|X = x]$:
  $$\hat{r}_M(X) = \int_{-\infty}^{\infty} r f_{R|X}(r|x) dr = \int_x^1 r \frac{1}{-r \ln x} dr = \frac{x-1}{\ln(x)}$$
  - That is, $\hat{R}_M(X) = \frac{X-1}{\ln(X)}$.

Estimation of a Random Variable: Case 3, Given a Random Variable (VI)

- In practice, $\hat{X}_M(Y) = E[X|Y]$ is impractical (hard to compute) for most uses.
- So, we resort to the linear estimator for X given Y (e.g., $\hat{x}_L(Y) = aY + b$).