CS70 - Lecture 26 Notes

Name: Felix Su SID: 25794773

Spring 2016 GSI: Gerald Zhang

Continuous Probability

> 1. Use **unions of intervals** to describe events

- Choose a real number X, uniformly at random in [0,L].
- Let [a,b] denote the event that the point X is in the interval [a,b].
  - $\Pr[[a,b]] = \frac{\text{length of } [a,b]}{\text{length of } [0,L]} = \frac{b-a}{L} = \frac{b-a}{1000}$
- Events in this space are unions of intervals.
- **Example:** the event A - within 50 of 0 or 1000 is
  - $A = [0,50] \cup [950,1000]$. Thus, $\Pr[A] = \Pr[[0,50]] + \Pr[[950,1000]] = \frac{1}{10}$

Finite vs. Continuous Probability Spaces

> 1. Start with probability of events (unions of intervals): $\Pr[A]$ for some events A
> 2. Probability is then a function from events to [0,1]
> 3. Function must be additive

- **Finite probability space:** $\Omega = \{1, 2, \ldots, N\}$
  - Started with probabilities of each outcome $\Pr[\omega] = p_\omega$
  - Defined probability of event is sum of probability of outcomes in the event: $\Pr[A] = \sum_{\omega \in A} p\omega$ for $A \subset \Omega$.
  - We used the same approach for countable Ω.
- **Continuous space:** $\Omega = [0, L]$,
  - Cannot start with $\Pr[\Omega]$, because this will typically be 0.
  - Start with probability of events (unions of intervals): $\Pr[A]$ for some events A. Here, we started with A = interval, or union of intervals.
  - Probability is then a function from events to [0,1]
  - Function must be additive. In our example, $\Pr[[0,50] \cup [950,1000]] = \Pr[[0,50]] + \Pr[[950,1000]]$

Example:
James Bond Shooting

> Chance of landing in a one foot radius circle that is inside a 4×5 rectangle.
> $\Omega = \{(x,y) : x \in [0,4], y \in [0,5]\}$.
> The size of the event is $\pi(1)^2 = \pi$.
> The "size" of the sample space which is 4×5.
> Since uniform, probability of event is $\frac{\pi}{20}$.

Continuous Random Variables: CDF

1. Define $\Pr[a < X \leq b] = \Pr[X \leq b] - \Pr[X \leq a] = F_X(b) - F_X(a)$

- Find function to define all intervals between a and b: $\Pr[a < X \leq b]$
- Cumulative probability Distribution Function of X (CDF of X) is
  - $F_X(x) = \Pr[X \leq x]$
- So, $\Pr[a < X \leq b] = \Pr[X \leq b] - \Pr[X \leq a] = F_X(b) - F_X(a)$.
  - Idea: two events $X \leq b$ and $X \leq a$.
  - Difference is the event $a < X \leq b$.
  - Indeed: $\{X \leq b\} - \{X \leq a\} = \{X \leq b\} \cap \{X > a\} = \{a < X \leq b\}$.

Cumulative Probability Distribution Function of X: CDF

$$F_X(x) = \Pr[X \leq x] \tag{1}$$

Define Probability of all Intervals

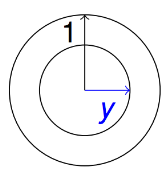
$$\Pr[a < X \leq b] = \Pr[X \leq b] - \Pr[X \leq a] = F_X(b) - F_X(a) \tag{2}$$

Example:
CDF: Value of X in [0,L] with L = 1000.

$$F_X(x) = \Pr[X \leq x] = \begin{cases} 0 & \text{for } x < 0 \\ \frac{x}{1000} & \text{for } 0 \leq x \leq 1000 \\ 1 & \text{for } x > 1000 \end{cases} \tag{3}$$

Probability that X is within 50 of center:
$\Pr[450 < X \leq 550] = \Pr[X \leq 550] - \Pr[X \leq 450] = \frac{550}{1000} - \frac{450}{1000} = \frac{100}{1000} = \frac{1}{10}$

Example:
CDF: Hitting random location on a unit circle.



Random Variable: Y distance from center.

Probability within y of center:

$$\Pr[Y \leq y] = \frac{\text{area of small circle}}{\text{area of dartboard}} = \frac{\pi y^2}{\pi} = y^2 \tag{4}$$

Hence,

$$F_Y(y) = \Pr[Y \leq y] = \begin{cases} 0 & \text{for } y < 0 \\ y^2 & \text{for } 0 \leq y \leq 1 \\ 1 & \text{for } y > 1 \end{cases} \tag{5}$$

Calculation Probability between .5 and .6 of center
$\Pr[0.5 < Y \leq 0.6] = \Pr[Y \leq 0.6] - \Pr[Y \leq 0.5] = F_Y(0.6) - F_Y(0.5) = .36 - .25 = .11$

Density function

1. Find probability of a certain value (within δ): $\lim_{\delta \to 0} \frac{\Pr[x < X \leq x+\delta]}{\delta} = \frac{d(F_X(x))}{dx}$

- Is the dart more like to be (near) .5 or .1?
- Probability within δ of x is $\Pr[x < X \leq x + \delta]$.
- Goes to 0 as δ goes to zero.
- Find the limit as δ goes to zero. $\lim_{\delta \to 0} \frac{\Pr[x < X \leq x+\delta]}{\delta}$
  * $= \lim_{\delta \to 0} \frac{\Pr[X \leq x+\delta] - \Pr[X \leq x]}{\delta}$
  * $= \lim_{\delta \to 0} \frac{F_X(x+\delta) - F_X(x)}{\delta}$
  * $= \frac{d(F_X(x))}{dx}$

Density

1. A **probability density function** for RV X with cdf $F_X(x) = \Pr[X \leq x]$ is the derivate of the cdf: $f_X(x) = \frac{d(F_X(x))}{dx}$
2. Probability that X is within δ of x, is $f_X(x)\delta$

- Definition: (Density) A probability density function for a random variable X with cdf $F_X(x) = \Pr[X \leq x]$ is the function $f_X(x)$ where:

$$F_X(x) = \int_{-\infty}^{x} f_X(u) du \tag{6}$$

- Thus, $\Pr[X \in (x, x+\delta]] = F_X(x+\delta) - F_X(x) \approx f_X(x)\delta$

Probability Density Function
For random variable X with cdf $F_X(x) = \Pr[X \leq x]$ is the function $f_X(x)$ where:

$$\Pr[X \in (x, x+\delta]] = F_X(x+\delta) - F_X(x) \approx f_X(x)\delta \tag{7}$$

Example:
Uniform over interval [0,1000]

$$f_X(x) = F_X(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{1000} & \text{for } 0 \leq x \leq 1000 \\ 0 & \text{for } x > 1000 \end{cases} \tag{8}$$

Example:
Uniform over interval [0,L]

$$f_X(x) = F_X(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{L} & \text{for } 0 \leq x \leq L \\ 0 & \text{for } x > L \end{cases} \tag{9}$$

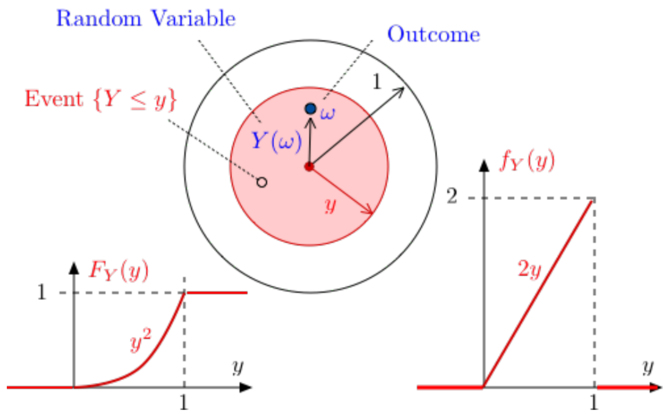
Example:
"Dart" board

- The cumulative distribution function (cdf) and probability distribution function (pdf) give full information.
- Use whichever is convenient.

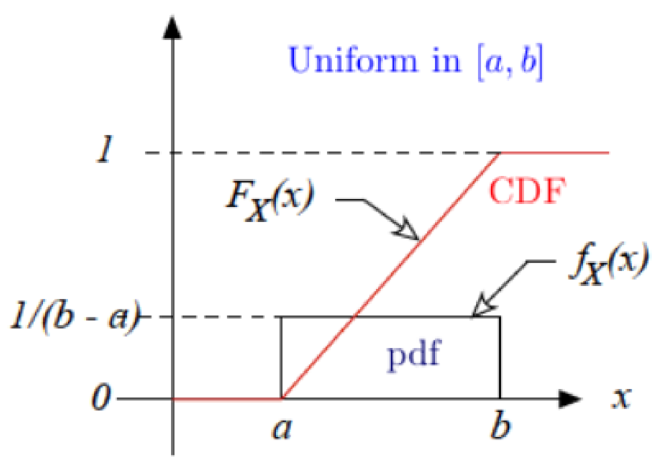
$$F_Y(y) = \Pr[Y \le y] = \begin{cases} 0 & \text{for } y < 0 \\ y^2 & \text{for } 0 \le y \le 1 \\ 1 & \text{for } y > 1 \end{cases} \tag{10}$$

$$f_Y(y) = F_Y'(y) = \begin{cases} 0 & \text{for } y < 0 \\ 2y & \text{for } 0 \le y \le 1 \\ 0 & \text{for } y > 1 \end{cases} \tag{11}$$

Target

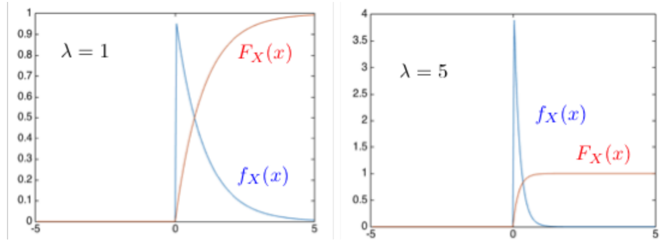


Uniform Distribution: U[a,b]



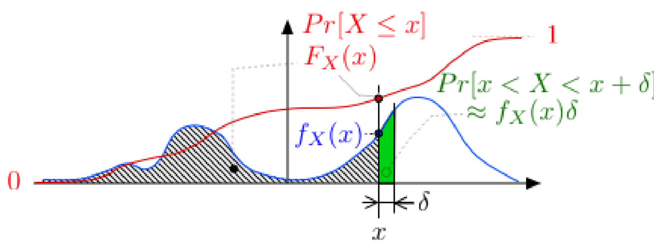
Expo(λ)

- Note that $\Pr[X > t] = e^{-\lambda t}$ for $t > 0$



The exponential distribution with parameter $\lambda > 0$ is defined by
$$f_X(x) = \lambda e^{-\lambda x} 1\{x \geq 0\} \tag{12}$$ $$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda x} & \text{if } x \geq 0 \end{cases} \tag{13}$$

Random Variables



Continuous random variable X, specified by

1. $F_X(x) = \Pr[X \leq x], \forall x$
   - **Cumulative Distribution Function (cdf)**: $\Pr[a < X \leq b] = F_X(b) - F_X(a)$
   - Non-decreasing between 0 and 1
     - $0 \leq F_X(x) \leq 1 \forall x \in \mathbb{R}$.
     - $F_X(x) \leq F_X(y)$ if $x \leq y$.
2. Or $f_X(x)$, where $F_X(x) = \int_{-\infty}^{x} f_X(u)du$ or $f_X(x) = \frac{d(F_X(x))}{dx}$
   - **Probability Density Function (pdf)**: $\Pr[a < X \leq b] = \int_a^b f_X(x)dx = F_X(b) - F_X(a)$
   - Non-negative and integrates to 1.
     - $f_X(x) \geq 0 \forall x \in \mathbb{R}$.
     - $\int_{-\infty}^{\infty} f_X(x)dx = 1$
3. Recall that $\Pr[X \in (x, x + \delta)] \approx f_X(x)\delta$.
   - Probability that you are δ away from x is $\approx f_X(x)\delta$
   - If density ($f_X(x)$) is large, more likely to be at x.
4. Think of X taking discrete values $n\delta$ for $n = \ldots, -2, -1, 0, 1, 2, \ldots$ with $\Pr[X = n\delta] = f_X(n\delta)\delta$

Some Examples

a. **Expo is memoryless**. Let X =Expo(λ). Then, for $s, t > 0$
   - $\Pr[X > t + s | X > s] = \frac{\Pr[X > t+s]}{\Pr[X > s]} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = \Pr[X > t]$.
   - 'Used is a good as new.'

b. **Scaling Expo**. Let X =Expo(λ) and $Y = aX$ for some $a > 0$. Then

- $\Pr[Y > t] = \Pr[aX > t] = \Pr[X > t/a] = e^{-\lambda(t/a)} = e^{-(\lambda/a)t} = \Pr[Z > t]$ for $Z =$Expo(λ/a)
- Thus, $a\times$ Expo(λ) = Expo(λ/a).

c. **Scaling Uniform** Let $X = U[0,1]$ and $Y = a + bX$ where $b > 0$. Then

- $\Pr[Y \in (y, y+\delta)] = \Pr[a + bX \in (y, y+\delta)] = \Pr[X \in (\frac{y-a}{b}, \frac{y+\delta-a}{b})]$
  - $= \Pr[X \in (\frac{y-a}{b}, \frac{y-a}{b} + \frac{\delta}{b})] = \frac{1}{b}\delta$, for $0 < \frac{y-a}{b} < 1$
  - $= \frac{1}{b}\delta$, for $a < y < a + b$.
- Thus, $f_Y(y) = \frac{1}{b}$ for $a < y < a + b$. Hence, $Y = U[a, a+b]$.

d. **Scaling pdf**. Let $f_X(x)$ be the pdf of X and $Y = a + bX$ where $b > 0$. Then

- $\Pr[Y \in (y, y+\delta)] = \Pr[a + bX \in (y, y+\delta)]$
  - $= \Pr[X \in (\frac{y-a}{b}, \frac{y+\delta-a}{b}]$
  - $= \Pr[X \in (\frac{y-a}{b}, \frac{y-a}{b} + \frac{\delta}{b}]$
  - $= f_X(\frac{y-a}{b})\frac{\delta}{b}$
- Now, the left-hand side is $f_Y(y)\delta$. Hence, $f_Y(y) = \frac{1}{b} f_X(\frac{y-a}{b})$.

Expectation

- **Definition** The expectation of a random variable X with pdf $f_X(x)$ is defined as $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$.
- Justification: Say $X = n\delta$ w.p. $f_X(n\delta)\delta$. Then,
  - $E[X] = \sum_n (n\delta)\Pr[X = n\delta] = \sum_n (n\delta) f_X(n\delta)\delta = \int_{-\infty}^{\infty} x f_X(x) dx$.
- Indeed, for any g, one has $\int g(x)dx \approx \sum_n g(n\delta)\delta$. Choose $g(x) = x f_X(x)$.

Expectation of function of RV

- **Definition** The expectation of a function of a random variable is defined as $E[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x) dx$.
- Justification: Say $X = n\delta$ w.p. $f_X(n\delta)\delta$. Then
  - $E[h(X)] = \sum_n h(n\delta)\Pr[X = n\delta]$
  - $= \sum_n h(n\delta) f_X(n\delta)\delta$
  - $= \int_{-\infty}^{\infty} h(x) f_X(x) dx$.
- Indeed, for any g, one has $\int g(x)dx \approx \sum_n g(n\delta)\delta$. Choose $g(x) = h(x) f_X(x)$.
- **Fact** Expectation is linear.
- **Proof** As in the discrete case.

$E[X]$ **with pdf** $f_X(x)$

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx \tag{14}$$

Exepection of a function of RV X**:** $E[h(X)]$ **with pdf** $f_X(x)$

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x) dx \tag{15}$$

Variance

- **Definition:** The variance of a continuous random variable X is defined as $\text{Var}[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - (\int_{-\infty}^{\infty} xf(x)dx)^2$

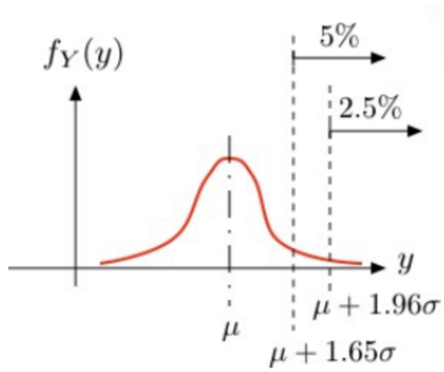
Variance

$$\text{Var}[X] = \int_{-\infty}^{\infty} x^2 f(x)dx - (\int_{-\infty}^{\infty} xf(x)dx)^2 \tag{16}$$

Motivation for Gaussian Distribution

- **Key fact:** The sum of many small independent RVs has a Gaussian distribution.
- This is the Central Limit Theorem. (See later.)
- Examples: Binomial and Poisson suitably scaled.
- This explains why the Gaussian distribution (the bell curve) shows up everywhere.

Normal Distribution



- For any mean: μ and std. dev: σ, a **normal** (aka **Gaussian**) random variable Y, which we write as $Y = \mathcal{N}(\mu, \sigma^2)$, has pdf $f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2}$
- Standard normal has $\mu = 0$ and $\sigma = 1$.
- Note: $\Pr[|Y - \mu| > 1.65\sigma] = 10\%$;$\Pr[|Y - \mu| > 2\sigma] = 5\%$.
- Guassian RV is within 2σ of the mean with 95%

Normal Distribution
For any μ and σ, a Gaussian RV, $Y = \mathcal{N}(\mu, \sigma^2)$ has pdf:

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2} \tag{17}$$

Summary: Continuous Probability

> 1. **pdf**: $\Pr[X \in (x, x + \delta]] = f_X(x)\delta$.
> 2. **CDF**: $\Pr[X \le x] = F_X(x) = \int_{-\infty}^{x} f_X(y)dy$.
> 3. **U[a,b], Expo(λ), target**
> 4. **Expectation**: $E[X] = \int_{-\infty}^{\infty} xf_X(x)dx$.

5. **Expectation of function**: $E[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x) dx$.

6. **Variance**: $\mathrm{Var}[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$.

7. **Gaussian**: $N(\mu, \sigma^2) : f_X(x) = \ldots$ "bell curve"