Section 6

continuous random variables :)

-Announcements-

PSet 4 due yesterday!

PSet 5 released, due **11:59pm Wednesday, August 5th**

Midterm: everyone gets free **+10** points!

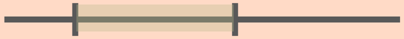
Discrete vs Continuous Random var.

discrete

the range consists of finite/countably infinite values

two "types" of random vars

continuous



the range consists of uncountably infinite values (*for example time is not discrete)*

two “types” of random vars

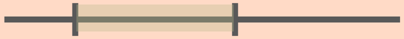
discrete

the range consists of finite/countably infinite values

PMF (prob. mass function)

$p_X(k) = P(X=k)$

continuous



the range consists of uncountably infinite values *(for example time is not discrete)*

discrete

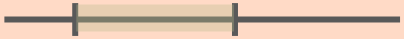
the range consists of finite/countably infinite values

PMF (prob. mass function)

$p_X(k) = P(X=k)$

two "types" of random vars

continuous



the range consists of uncountably infinite values (*for example time is not discrete)*

PMF (prob. mass function)

$p_X(k) = P(X=k) = 0$

two "types" of random vars

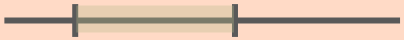
discrete

the range consists of finite/countably infinite values

PMF (prob. mass function)

$p_X(k) = P(X=k)$

continuous



the range consists of uncountably infinite values (*for example time is not discrete*)

*PDF (prob. **density** function)*

$f_X(k)$!= $P(X=k)$

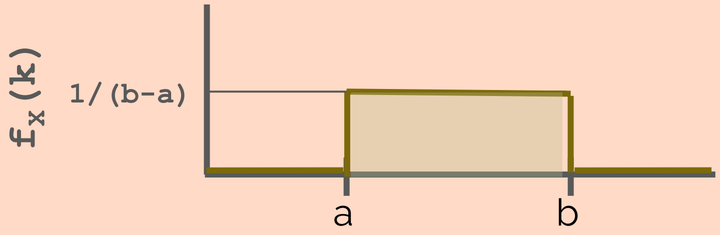
discrete vs. continuous

	Discrete	Continuous
PMF/PDF	$p_X(x) = P(X = x)$	$f_X(x) \neq P(X = x) = 0$
CDF	$F_X(x) = \sum_{t \leq x} p_X(t)$	$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$
Normalization	$\sum_{x} p_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
Expectation	$\mathbb{E}[g(X)] = \sum_{x} g(x)\,p_X(x)$	$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)\,f_X(x)\,dx$

Zoo of continuous RVs

Uniform RV (continuous version)

X~Unif(a, b) randomly takes on any real number between a and b



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

$$\mathbb{E}[X] = \frac{a+b}{2}$$

$$\mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

Exponential RV

X~Exp(λ) tells how much time till a certain event happens
(λ *is the rate of time*)

think of this as the "continuous version" of the geometric distribution!

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}[X] = \frac{1}{\lambda}$$

$$\mathrm{Var}(X) = \frac{1}{\lambda^2}$$

don't confuse this with the Poisson distribution just bc it's related with time, they're very different!
(Poisson is *number* of events in a certain time frame)

$$F_X(x) = 1 - e^{-\lambda x}$$

$F_X(x)$ = P(X <= x) *this is the integral of* $f_X(x)$

Normal/Gaussian RV

X~N(μ, σ^2)
Standard normal is Z ~ N(0, 1)

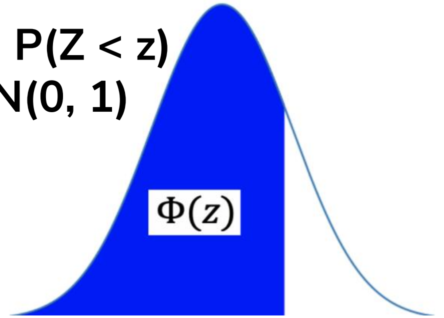
general strategy:

1. **Find** μ and σ^2 of normal RV X
2. **Standardize** $X \sim N(\mu, \sigma^2)$ to get $Z = (X-\mu)/\sigma \sim N(0,1)$
3. Use **Phi table** to get appropriate value with $Z = (X-\mu)/\sigma$
4. **Solve** for X

PDF:

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

$\Phi(z) = P(Z < z)$
if $Z \sim N(0, 1)$



Closure of the Normal Distribution

Closure for Normal Distribution

Let $X \sim N(\mu, \sigma^2)$. Then $aX + b \sim N(a\mu + b, b^2\sigma^2)$

More closure for normal

Let X1,..,Xn be <u>independent</u> normal RVs with $E[X_i]=\mu_i$ and $Var(X_i=\sigma_i^2)$.

$$X = \sum_{i=1}^{n}(a_i X_i + b) \sim \mathcal{N}\left(\sum_{i=1}^{n}(a_i\mu_i + b), \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$$

Central Limit Theorem

Theorem. (Central Limit Theorem) $X_1, \ldots, X_n$ i.i.d. with mean μ and variance σ^2. Let $Y_n = \frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$. Then,

$$\lim_{n\to\infty} Y_n \to \mathcal{N}(0,1)$$

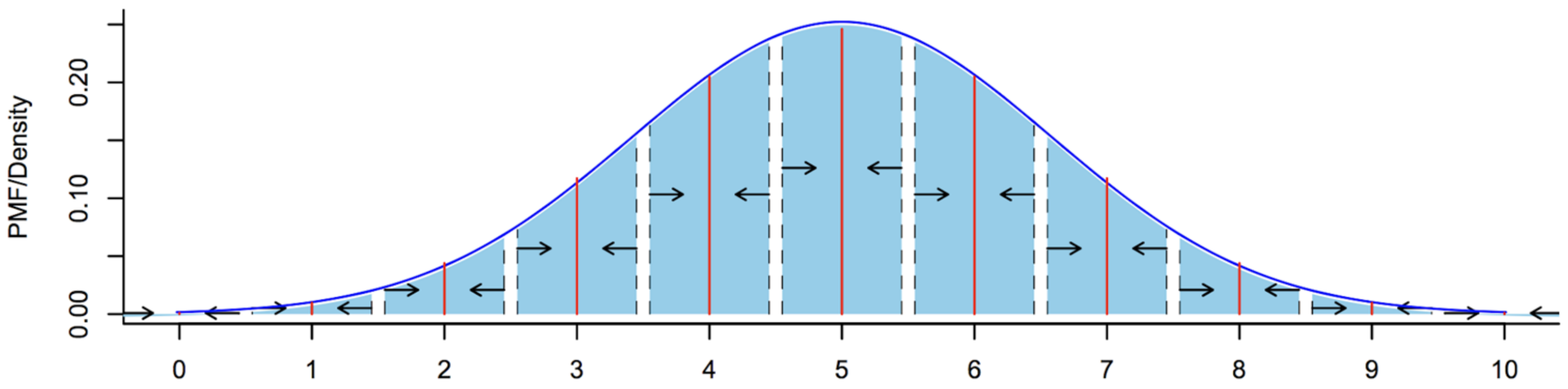
One main application:

Use Normal Distribution to Approximate Y_n

No need to understand Y_n !!

Solution – Continuity Correction

Probability estimate for i: Probability for all x that round to i!



To estimate probability that discrete RV lands in (integer) interval $\{a, \ldots, b\}$, compute probability continuous approximation lands in interval $[a - \frac{1}{2}, b + \frac{1}{2}]$

Content Review (Problem #1)

Problem 1a

a) What is $\mathbb{P}(X = 4)$ if X is a **continuous** random variable?

- ○ 1
- ○ 0
- ○ not enough information

If X is a continuous random variable, the probability that it takes on a particular constant is 0 since the support of X has infinite real values.

Problem 1b

b) The cumulative distribution function for a continuous random variable X is $F_X(k) =$

- ○ $\int_{-\infty}^{k} f_X(x)dx$
- ○ $\int_{-\infty}^{\infty} f_X(x)dx$
- ○ $\int_{k}^{\infty} f_X(x)dx$
- ○ $\frac{d}{dk} f_X(k)$

We take the integral over the PDF over the appropriate range to get the CDF. Since the CDF is F(*k*) = P(*X* <= *k*), we take the integral from negative infinity up to *k*.

Problem 1c

c) The probability density function for a continuous random variable X is $f_X(k) =$

- ○ $\int_{-\infty}^{k} f_X(x)dx$
- ○ $\frac{d}{dk}F_X(k)$

We take the derivative of the CDF to get the PDF.

Problem 1d

d) True or **False**. If X is a continuous random variable, $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx$

True. This is by the definition of expectation for continuous random variables. Note the difference from the discrete case is that we're using an integral instead of a summation, and we're using density instead of pmf!

Problem 1e

e) **True** or **False**. If X is a continuous random variable, $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

True. This definition for variance applies regardless of whether X is discrete or continuous.

Problem 1f

f) Which of the following follow an $\mathsf{Exponential}(\lambda)$ distribution?

- ○ Number of minutes to the first success with λ as average number of successes per minute
- ○ Number of successes in the first 1 minute with λ as average number of successes per minute
- ○ Time (real number) to the first success with λ as average number of successes per minute

The exponential random variable is from our zoo of continuous random variables, and represents the time (continuous) till the first success. Note that (b) is a Poisson random variable with parameter λ!

Problem 1g

g) True or False: For any random variable X, $\mathbb{P}(X = 5) = \mathbb{P}(X - 5 = 0)$.

True. We can think of *X* - 5 as another random variable where we take the output of *X* and subtract 5 from it. Then the probability that *X* - 5 is zero is identical to the probability that *X* is originally five.

Problem 1h

h) True or False: For some continuous random variable X, $\mathbb{P}(X \leq 5) \neq \mathbb{P}(X < 5)$.

False. Note that P(X <= 5) = P(X = 5) + P(X < 5). But the first term is zero, so the probabilities are exactly equal. This holds for every continuous random variable.

Problem 1i

i) True or False: Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $a, b \in \mathbb{R}$. Then $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

True. This follows by the closure of the normal distribution.

Problem 7

Problem 7 - Throwing a dart

Consider the closed unit circle of radius r, i.e., $S = \{(x, y) : x^2 + y^2 \leq r^2\}$. Suppose we throw a dart onto this circle and are guaranteed to hit it, but the dart is equally likely to land anywhere in S. Concretely this means that the probability that the dart lands in any particular area of size A (that is entirely inside the circle of radius R), is equal to $\frac{\text{A}}{\text{Area of whole circle}}$. The density outside the circle of radius r is 0.

Let X be the distance the dart lands from the center. What is the CDF and pdf of X? What is $\mathbb{E}[X]$ and $\text{Var}(X)$?

Problem 9

Problem 9 - Will the battery last?

Suppose that the number of miles that a car can run before its battery wears out is exponentially distributed with expectation 10,000 miles. If the owner wants to take a 5000 mile road trip, what is the probability that she will be able to complete the trip without replacing the battery, given that the car has already been used for 2000 miles on the road trip?

Problem 12

Problem 12 - Normal Questions

a) Let X be a normal random variable with parameters $\mu = 10$ and $\sigma^2 = 36$. Compute $\mathbb{P}(4 < X < 16)$.

b) Let X be a normal random variable with mean 5. If $\mathbb{P}(X > 9) = 0.2$, approximately what is $\mathrm{Var}(X)$?

c) Let X be a normal random variable with mean 12 and variance 4. Find the value of c such that

$$\mathbb{P}(X > c) = 0.10.$$

Problem 20

Problem 20 - Tweets

A prolific twitter user tweets approximately 350 tweets per week. Let's assume for simplicity that the tweets are independent, and each consists of a uniformly random number of characters between 10 and 140. (Note that this is a discrete uniform distribution.) Thus, the central limit theorem (CLT) implies that the number of characters tweeted by this user is approximately normal with an appropriate mean and variance. Assuming this normal approximation is correct, estimate the probability that this user tweets between 26,000 and 27,000 characters in a particular week. (This is a case where continuity correction will make virtually no difference in the answer, but you should still use it to get into the practice!).

Thank you!