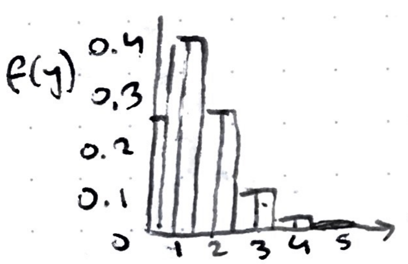
4/25/19 Lecture 8

New due date for Take-Home Test 1: May 5 by 11:59pm

DD Office Hours Next Week: T&Th 3:30-5pm in Jack's Lounge
MWF are still to be decided

Def: Given a discrete rv Y, the probability mass function (pmf or pf) of Y is the function f_Y that keep track of the probabilities associated with Y: $f_Y(y) = P(Y=y)$

The set $\{y : f_Y(y) > 0\}$ is called the support of (the distribution of) Y



pf of Y in the T-S case study

← Histogram

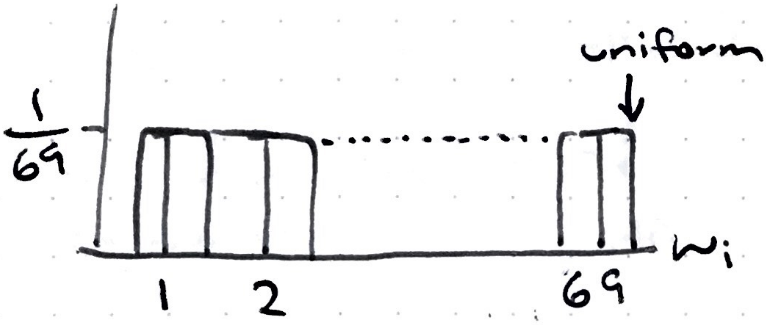
Def: A rv Y that only takes on the values $\{0, 1\}$ - ie a binary rv is a Bernoulli distribution with parameter p

Ex: Take Home Test Problem 2: 5 balls are drawn at random without replacement from a bin with balls numbered $\{1, 2, \ldots, 69\}$

Let W_i = # on ith drawn white ball $(i = 1, \ldots, 5)$

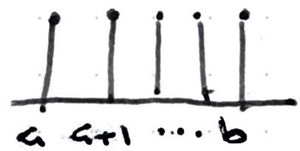
Clearly, $P(W_1 = w_1) = \begin{cases} \frac{1}{69} & \text{for } w_1 = 1 \text{ or } 2 \text{ or} \ldots \text{or } 69 \\ 0 & \text{otherwise} \end{cases}$

Less clearly, but true, $W_2 \ldots W_5$ follow the same distribution if nothing is known about the previous draws



$$\sum_{y=a}^{b} \overset{\geq 0}{f_{Y}(y)} = 1 \quad = \sum_{y=a}^{b} \left(\frac{1}{b-a+1} \right) = 1$$

$Y \sim \text{Uniform}\{a, b\}$



Def: For any 2 integers $a \le b$ a rv Y that's equally likely to be any of the values $\{a, a+1, \dots, b\}$ has the same uniform distribution uniform$\{a, b\}$

Evidently its pf is $f_Y(y) = P(Y=y) = \begin{cases} \frac{1}{b-a+1} & \text{for } y=a,\dots,b \\ 0 & \text{else} \end{cases}$

"is distributed as"

$Y \sim \text{uniform}\{a, b\} \Leftrightarrow Y$ chosen at random from $\{a, a+1, \dots, b\}$

Def: n random trials are performed with each trial recorded as a success S or failure F. If each trial is independent of all the others and the chance of success is constant across the trials then $Y = \#$ of successes has the binomial distribution

$$f_Y(y) = P(Y=y) = \begin{cases} \binom{n}{y} p^y (1-p)^{n-y} & \text{for } y=0,1,\dots,n \\ 0 & \text{else} \end{cases}$$

pf

with parameters

$Y \sim \text{Binomial}(n, p)$ or $(Y | n, p)$

Bernoulli distribution is written as Bernoulli(p)

if $f_Y(y) = P(Y=y) = \begin{cases} p & \text{for } y=1 \\ 1-p & y=0 \\ 0 & \text{else} \end{cases}$

$= p^y (1-p)^{1-y}$

Notation: Y follows a Bernoulli distribution

$Y \sim \text{Bernoulli}(p)$ or $(Y | p)$

In shorthand $Y \sim \text{Binomial}(n, p)$ or $(Y | n, p)$

Let $B_i = \begin{cases} 1 & \text{if trial } i \text{ is a success} \\ 0 & \text{if trial } i \text{ is a failure} \end{cases}$ for $i=1,\ldots,n$

then under these assumptions $B_i \overset{IID}{\sim}$ Bernoulli(p) and all the B_i are independent

Notation: $X_i \overset{IID}{\sim} f_X(x_i)$ means that all of the rvs

$X_1, X_2, \ldots$ are independent and identically distributed draws from the distribution with pf $f_{X_i}(x_i)$

Thus with the success/failure trials, $B_i \overset{IID}{\sim}$ Bernoulli(p) and $(Y = \sum_{i=1}^{n} B_i)\ (i=1,\ldots,n) \sim$ Binomial(n,p)

This is our 1st example of the distribution of the sum of a bunch of IID rvs

Case Study

The Poisson Study (p. 294)

You work for a bank that has lately been receiving complaints about long waiting times in the bank teller lines. To quantify the extent of the problem you gather a data set like the one below on Mon–Sat in a randomly chosen week. This is a new type of data set that unfolds in time. There are 2 equivalent ways to keep track of the arrivals: you could let $N(t)$ = # arrivals in $[0,t]$ or you could keep track of the inter-arrival times (times between arrivals) $T_1, T_2, \ldots$

time	event
9 am	bank opens
9:01	arrival
9:03	arrival
9:04	arrival
9:05	departure

Here we'll look at $N(t)$.

Def: A stochastic process is a collection of random variables indexed by elements of an indexing set, usually $t \in T$, where t represents time.

9am ex: 5pm

# of arrivals in $[0, t] \to N(t)$, $t \in [0, t_{max}]$ is an example of a stochastic process, in which the time index is continuous; $\{T_1, T_2, \ldots\}$ (interarrival times) is an example of a stochastic process with a discrete time index.

$(T_t, t = 1, 2, \ldots)$

What may reasonably be assumed about the random behavior of $N(t)$?

Assumptions:

nonoverlapping

1) The #s of arrivals in any collection of disjoint time intervals are (mutually) independent (reasonably if unrelated customers arrive at the bank haphazardly in time)

2) P(1 or more arrivals in short time interval $[s, s+t]$, $t > 0$ small) is proportional to t, for ex $\lambda \cdot t$ for a rate parameter $\lambda > 0$ (this is reasonable if the arrivals process is smooth rather than lumpy)

Def: To say that a function $f(t)$ of a small t is $o(t)$ is to say that $\lim_{t \to 0} \frac{f(t)}{t} = 0$. In other words, $f(t)$ approaches 0 as $t \to 0$ at a rate faster than t itself

The mathematically formal way to state Assumption 2 is then $P(\text{at least one arrival in } [s, s+t]) = \lambda t + o(t)$

Note that assumption 2 implies that the rate parameter λ is constant in time; this would be unrealistic in the bank problem over an entire day but would be reasonable during stable subsets of a day (e.g. 10am - 11:30 am)

3) nearly simultaneous arrivals are rare

$$P(2 \text{ or more arrivals in } [s, s+t]) = o(t)$$

Remarkably, these 3 simple and often plausible assumptions specify the probability behavior of $N(t)$ uniquely (see exercise 16 pg. 296 for a proof of the following result)

λ constant $\leftrightarrow$ $N(t)$ is a stationary stochastic process

Under assumptions 1-3 above, $N(t)$ is a Poisson process with rate parameter λ

<u>Exploring the Poisson distribution</u>

$$f(y \mid \lambda) = p(Y = y \mid \lambda) = \begin{cases} \dfrac{\lambda^y e^{-\lambda}}{y!} & y = 0, 1, \ldots \\ 0 & \text{else} \end{cases}$$

Def: For any $\lambda > 0$ a random variable Y has the Poisson distribution with parameter λ (Poisson(λ)) if its pf is
$$f_Y(y|\lambda) = P(Y=y|\lambda) = \begin{cases} \dfrac{\lambda^y e^{-\lambda}}{y!} & \text{for } y=0,1,\ldots \\ 0 & \text{else} \end{cases}$$

Result: under assumptions 1-3, if Y = (# of arrivals in any time interval of length) then $Y \sim$ Poisson (λt)

Def: A poisson process with rate λ per unit time is a stochastic process satisfying

1) for any $s > 0$, # of arrivals in $[s, s+t] \sim$ Poisson (λt)

2) # of arrivals in disjoint time intervals are independent

Continuous random variables

Ex: Round-off error in CS

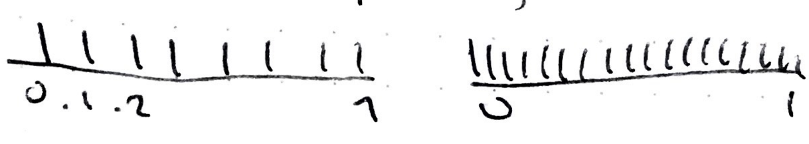
single-precision floating point #s carry about 7 decimal sigfigs of accuracy, leading to roundoff error in the last digit

$$3.141592|653589$$
$$\uparrow$$
$$3 \text{ causing } .04 \text{ error}$$

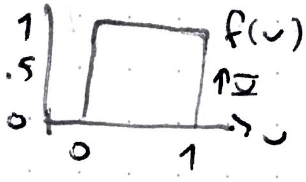
It's important to study how these errors accumulate as the # of steps

Since there's no reason one decimal digit would be favored over another in rounding, the uniform distribution is key to these calculations

Consider 1st uniform $\{0, 0.1, \ldots 0.9\}$ and then $\{0, 0.01, 0.02, \ldots, 0.99\}$



In the limit with more and more sigfigs this should go to



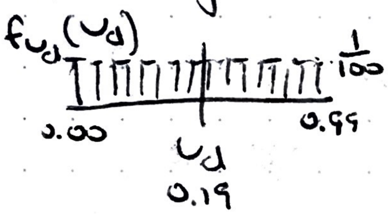
the continuous uniform distribution uniform (0, 1) on the unit interval

has to be square ↗

The analogue of the discrete pf in this case is the smooth continuous function

$$f_{\underline{U}}(u) = \begin{cases} 1 & \text{for } 0 \leq u \leq 1 \\ 0 & \text{else} \end{cases}$$

* analogue of summation is integration



U_d $P(0.0 \leq u_d \leq 0.19)$

$$= \sum_{u_d = 0.00}^{0.19} f_{\underline{U}_d}(u_d) = 0.2$$

<u>Def</u>: A random variable ($\underline{Y}$ has a continuous distribution) $\leftrightarrow$ ($\underline{Y}$ is a continuous rv) if there exists a continuous non-negative function $f_{\underline{Y}}$ defined on $\mathbb{R}$ such that for every interval $[a, b]$, $P(a \leq \underline{Y} \leq b) =$

$$\int_a^b f_{\underline{Y}}(y)\, dy$$

In this definition, a can be $-\infty$ and b can be too.

<u>Def</u>: If $\underline{Y}$ is a continuous rv, the function $f_{\underline{Y}}$ in the previous definition is called the <u>probability density function</u> (PDF or pdf) of $\underline{Y}$.

The set $y = \{ y : f_{\underline{Y}}(y) > 0 \}$ is called the support of (the distribution of) $\underline{Y}$

Clearly (a) $f_{\underline{Y}}(y) \geq 0$ for all y

(b) $\int_{-\infty}^{\infty} f_{\underline{Y}}(y)\, dy = 1$

What about individual points — singletons — $\{y\}$ on $\mathbb{R}$?

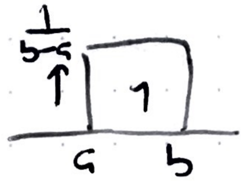
If $f_{\underline{Y}}$ is continuous on its support $\int_a^b f_{\underline{Y}}(y)\,dy$ can equally well stand for $P(a \le \underline{Y} \le b)$ or $P(a < \underline{Y} \le b)$ or $P(a \le \underline{Y} < b)$ or $P(a < \underline{Y} < b)$, because (e.g.) $\int_a^a f_{\underline{Y}}(y)\,dy = 0$ if $f_{\underline{Y}}$ is continuous at $y = a$.

Thus, importantly, $\boxed{P(\underline{Y} = y) = 0 \text{ for all } -\infty < y < \infty}$

Weirdly, this doesn't mean that the value y of $\underline{Y}$ is impossible, as it does with discrete rv; it just means that singletons have to have 0 probability (otherwise $\int_{-\infty}^{\infty} f_{\underline{Y}}(y)\,dy = +\infty$ not 1)

<u>Def</u>: With any two real #s a and b satisfying $a < b$, $\underline{Y} \sim \text{Uniform}(a, b) \overset{\text{iff}}{\longleftrightarrow} P(\underline{Y}$ is in any subinterval of $[a, b])$ = the length of the subinterval $\overset{\text{iff}}{\longleftrightarrow}$

$$f_{\underline{Y}}(y) = \left\{ \begin{array}{ll} \frac{1}{b-a} & \text{for } a \le y \le b \\ 0 & \text{else} \end{array} \right\}$$



<u>Def</u>: The indicator function for any (true/false) proposition A is $I(A) = \begin{cases} 1 & \text{if } A \text{ true} \\ 0 & \text{if } A \text{ false} \end{cases}$

$$I_A(y) = \begin{cases} 1 & \text{if } y \in A \\ 0 & \text{else} \end{cases}$$

With this definition, $Y \sim \text{Uniform}(a,b) \longleftrightarrow$

$$f_Y(y) = \frac{1}{b-a} I(a \le y \le b) = \frac{I_{[a,b]}(y)}{b-a}$$

<u>Contrast</u>

$Y \sim \text{Uniform}(a,b)$ continuous and uniform on (a,b) or $(a,b]$ or $[a,b]$ or $[a,b)$

$Y \sim \text{Uniform}\{a,b\}$ for a,b integers with $a<b$ $\longleftrightarrow$ Y discrete and uniform on $\{a, a+1, \dots, b\}$

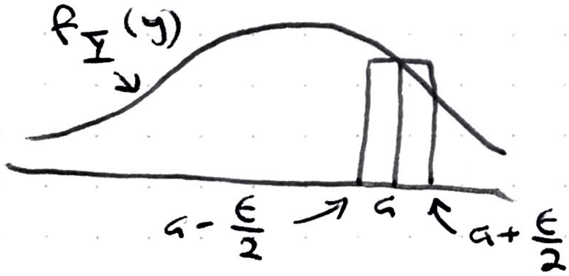
Density values $f_Y(y)$ are themselves not probabilities. For ex, they can easily be >1 and can even be $+\infty$

Density values define probabilities

$$P(a \le Y \le b) = \int_a^b f_Y(y)\,dy$$

For small $\epsilon > 0$ you can see from their sketch that

$$P\left(a - \frac{\epsilon}{2} \le Y \le a + \frac{\epsilon}{2}\right) = \int_{a-\frac{\epsilon}{2}}^{a+\frac{\epsilon}{2}} f_Y(y)\,dy = \text{area of rectangle} = \epsilon \cdot f_Y(a)$$



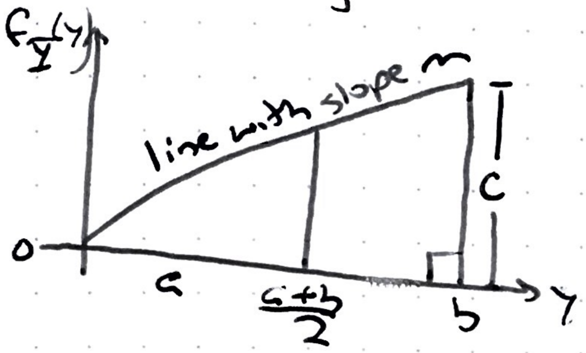
(connection with histograms)

$$\underbrace{f_Y(a)}_{\text{density}} = \frac{P\left(a - \frac{\epsilon}{2} \le Y \le a + \frac{\epsilon}{2}\right)}{\epsilon} \begin{array}{l} \}\text{ probabilities} \\ \}\text{ horizontal slice} \end{array}$$

(probability concentration near a)

Can a continuous rv Y have a pdf that looks like a triangle?

Ex: a triangular distribution



$m = \frac{c}{b-a}$

What has to be true about a pdf?

- ☑ It's always ≥ 0
- ☑ its integral from $-\infty$ to $+\infty$ is 1

The line has slope $\frac{c}{b-a} = m$ and passes through the point $(x, y) = (a, 0)$ so the equation of the line is

$$y - y_1 = m(x - x_1) \longleftrightarrow y = \frac{c}{b-a}(x-a) = \cancel{f_Y(x)}$$

y $\quad$ $f_Y(y)$

axis notational collision

Densities have to integrate to 1 so $\int_a^b \frac{c}{b-a}(x-a)\,dx = 1$

$\longleftrightarrow c = \frac{2}{b-a}$

Easier way: Area of a triangle is $\frac{\text{base} \cdot \text{height}}{2}$ so

$1 = \frac{1}{2}(b-a)c$ and $c = \frac{2}{b-a}$