Probability Review

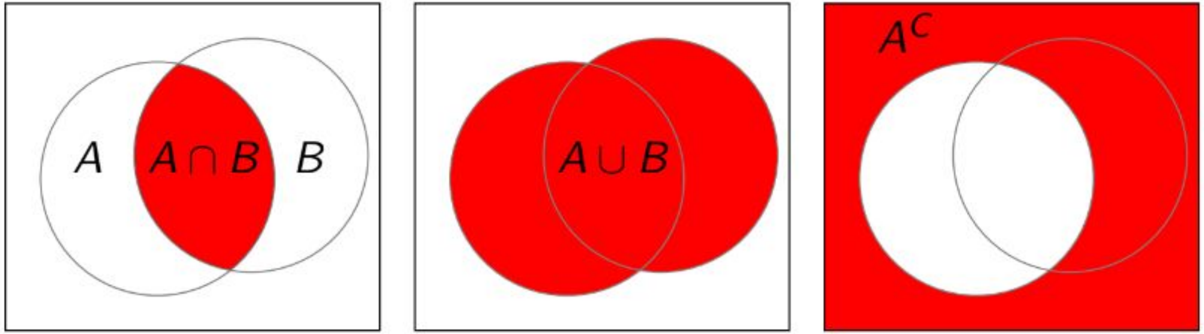
18662: Recitation 1

<u>Sources:</u>
10-701 Spring 2013 Machine Learning, CMU
Spring 2018, Introduction To Probability, MIT OCW
2010, Rob Hall, CMU
18751, Fall 2023, CMU

Set Theory

A *set* is just a collection of *elements* denoted e.g.,
$S = \{s_1, s_2, s_3\}, R = \{r : \text{some condition holds on } r\}$.

- **Intersection**: the elements that are in both sets:
  $A \cap B = \{x : x \in A \text{ and } x \in B\}$
- **Union**: the elements that are in either set, or both:
  $A \cup B = \{x : x \in A \text{ or } x \in B\}$
- **Complementation**: all the elements that aren't in the set:
  $A^C = \{x : x \notin A\}$.



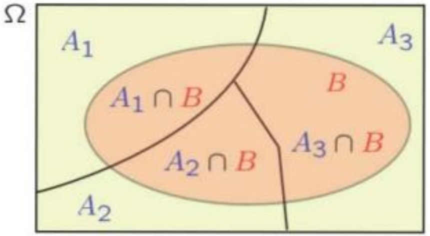
Properties

- **Commutativity**: $A \cup B = B \cup A$
- **Associativity**: $A \cup (B \cup C) = (A \cup B) \cup C$.
- **Distributive properties**: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Mutually Exclusive

- A sequence of sets $A_1, A_2 \ldots$ is called **pairwise disjoint** or **mutually exclusive** if for all $i \neq j$, $A_i \cap A_j = \{\}$.
- If the sequence is pairwise disjoint and $\bigcup_{i=1}^{\infty} A_i = S$, then the sequence forms a **partition** of S.



Partitions are useful in probability theory and in life:

$$
\begin{aligned}
B \cap S &= B \cap \left(\bigcup_{i=1}^{\infty} A_i\right) \quad \text{(def of partition)} \\
&= \bigcup_{i=1}^{\infty} (B \cap A_i) \quad \text{(distributive property)}
\end{aligned}
$$

Note that the sets $B \cap A_i$ are also pairwise disjoint

Terminology

Name	What it is	Common Symbols	What it means
Sample Space	Set	Ω, S	"Possible outcomes."
Event Space	Collection of subsets	$\mathcal{F}, E$	"The things that have probabilities.."
Probability Measure	Measure	P, π	Assigns probabilities to events.
Probability Space	A triple	$(\Omega, \mathcal{F}, P)$	

$$\Omega = \{1,2,3,4,5,6\}$$
$$\mathcal{F} = 2^{\Omega} = \{\{1\},\{2\}\ldots\{1,2\}\ldots\{1,2,3\}\ldots\{1,2,3,4,5,6\},\{\}\}$$

$P(\{1\}) = P(\{2\}) = \ldots = \frac{1}{6}$ (i.e., a fair die)
$P(\{1,3,5\}) = \frac{1}{2}$ (i.e., half chance of odd result)
$P(\{1,2,3,4,5,6\}) = 1$ (i.e., result is "almost surely" one of the faces).

Axioms of probability

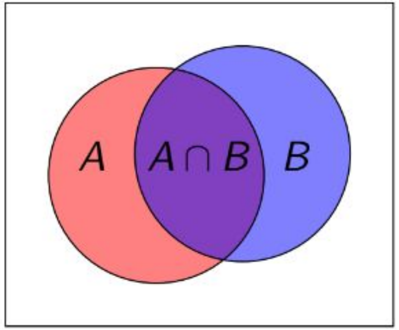
A set of conditions imposed on probability measures (due to Kolmogorov)

- $P(A) \geq 0, \forall A \in \mathcal{F}$
- $P(\Omega) = 1$
- $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ where $\{A_i\}_{i=1}^{\infty} \in \mathcal{F}$ are pairwise disjoint.

These quickly lead to:

- $P(A^C) = 1 - P(A)$ (since $P(A) + P(A^C) = P(A \cup A^C) = P(\Omega) = 1$).
- $P(A) \leq 1$ (since $P(A^C) \geq 0$).

Conditional probability



For events $A, B \in \mathcal{F}$ with $P(B) > 0$, we may write the **conditional probability of A given B**:

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

Interpretation: the outcome is definitely in B, so treat B as the entire sample space and find the probability that the outcome is also in A.

This rapidly leads to: $P(A|B)P(B) = P(A \cap B)$ aka the "chain rule for probabilities." (why?)

When $A_1, A_2 \ldots$ are a partition of Ω:

$$P(B) = \sum_{i=1}^{\infty} P(B \cap A_i) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i)$$

This is also referred to as the "law of total probability."

Bayes' Rule

Using the chain rule we may see:

$$P(A|B)P(B) = P(A \cap B) = P(B|A)P(A)$$

Rearranging this yields **Bayes' rule**:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Often this is written as:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_i P(A|B_i)P(B_i)}$$

Where B_i are a partition of Ω (note the bottom is just the law of total probability).

Example problem

A person uses his car 30% of the time, walks 30% of the time and rides the bus 40% of the time as he goes to work. He is late 10% of the time when he walks; he is late 3% of the time when he drives; and he is late 7% of the time he takes the bus.

a. What is the probability he took the bus if he was late?
b. What is the probability he walked if he is on time?

Solution

a. $P(B|L) = \dfrac{(0.40)(0.07)}{(0.40)(0.07) + (0.30)(0.03) + (0.30)(0.10)} \doteq 0.418$

b. $P(W|T) = \dfrac{(0.30)(.90)}{(0.30)(0.97) + (0.30)(0.90) + (0.40)(0.93)} \doteq 0.289$

Example problem 2

Suppose we have 3 cards identical in form except that both sides of the first card are colored red, both sides of the second card are colored black, and one side of the third card is colored red and the other side is colored black.

The 3 cards are mixed up in a hat, and 1 card is randomly selected and put down on the ground.

If the upper side of the chosen card is colored red, what is the probability that the other side is colored black?

Solution

Let RR, BB, and RB denote, respectively, the events that the chosen cars is the red-red, the black-black, or the red-black card. Letting R be the event that the upturned side of the chosen card is red, we have that the desired probability is obtained by

$$
\begin{aligned}
P(RB \mid R) &= \frac{P(RB \cap R)}{P(R)} \\
&= \frac{P(R \mid RB)P(RB)}{P(R \mid RR)P(RR) + P(R \mid RB)P(RB) + P(R \mid BB)P(BB)} \\
&= \frac{(\frac{1}{2})(\frac{1}{3})}{(1)(\frac{1}{3}) + (\frac{1}{2})(\frac{1}{3}) + 0(\frac{1}{3})} = \frac{1}{3}
\end{aligned}
$$

This question was actually just like the Monty Hall problem!

Independence

Two events A, B are called **independent** if $P(A \cap B) = P(A)P(B)$.

When $P(A) > 0$ this may be written $P(B|A) = P(B)$
e.g., rolling two dice, flipping n coins etc.

Two events A, B are called **conditionally independent given** C when $P(A \cap B|C) = P(A|C)P(B|C)$.

When $P(A) > 0$ we may write $P(B|A, C) = P(B|C)$
e.g., "the weather tomorrow is independent of the weather yesterday, knowing the weather today."

Information on some of the events does not change probabilities related to the remaining events

Example problem 3

Jake is shooting free throws. Making or missing free throws doesn't change the probability that he will make his next one, and he makes his free throws 88% of the time.

What is the probability of Jake making all of his next 9 free throw attempts?

Is independence the same as mutually exclusive?

Independence: Two events are independent if the occurrence of one event does not affect the probability of the other event occurring. In other words, the events have no influence on each other.

Example: Consider the act of flipping a coin and rolling a die. Flipping a coin (resulting in heads or tails) has no effect on the outcome of rolling a die (resulting in a number from 1 to 6), and vice versa. These two events are independent.

Mutual Exclusivity: Two events are mutually exclusive if they cannot occur at the same time. In other words, the occurrence of one event means the other cannot possibly happen.

Example: Consider drawing a single card from a standard deck of 52 cards. The event of drawing a heart and the event of drawing a club are mutually exclusive. If you draw a heart, it cannot simultaneously be a club, and vice versa.

An example of events that are independent but not mutually exclusive involves rolling two dice. Let's define two events:

Event A: Rolling an even number on the first die.

Event B: Rolling an even number on the second die.

These events are independent because the outcome of the first die does not affect the outcome of the second die. The probability of rolling an even number on one die is not influenced by the result of the other die.

However, these events are not mutually exclusive. Mutually exclusive events cannot happen at the same time, but in this case, both Event A and Event B can occur simultaneously. For example, if you roll a 4 on the first die (Event A occurs) and a 6 on the second die (Event B occurs), both events have happened together.

Thus, this scenario illustrates two events that are independent (the outcome of one die doesn't affect the outcome of the other) but not mutually exclusive (both can happen at the same time).

An example of mutually exclusive events that are not independent involves a single card draw from a standard deck of 52 playing cards. Let's define two events:

Event A: Drawing an Ace.

Event B: Drawing a King.

These events are mutually exclusive because a single card cannot be both an Ace and a King at the same time. If you draw an Ace, it is impossible for that same card to be a King, and vice versa.

However, these events are not independent. Independence implies that the occurrence of one event does not affect the probability of the other. In this case, if you know that Event A has occurred (you've drawn an Ace), the probability of Event B occurring (drawing a King) becomes 0, because the card drawn is already known to be an Ace. Similarly, if Event B occurs, the probability of Event A is 0. This dependence means the events cannot be independent.

So, in this scenario, you have two events (drawing an Ace and drawing a King) that are mutually exclusive (cannot occur together) but not independent (the occurrence of one affects the probability of the other).

Random Variable

A **random variable** is a function $X : \Omega \to \mathbb{R}^d$

associates a value to the outcome of a randomized event

e.g.,

- Roll some dice, $X =$ sum of the numbers.
- Indicators of events: $X(\omega) = 1_A(\omega)$. e.g., toss a coin, $X = 1$ if it came up heads, 0 otherwise.
- Give a few monkeys a typewriter, $X =$ fraction of overlap with complete works of Shakespeare.
- Throw a dart at a board, $X \in \mathbb{R}^2$ are the coordinates which are hit.

Cumulative Distribution Function

- $F_X(x) = P(X \leq x)\ \forall x \in \mathcal{X}$
- The CDF completely determines the probability distribution of an RV
- The function $F(x)$ is a CDF i.i.f
  - $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
  - $F(x)$ is a non-decreasing function of x
  - $F(x)$ is right continuous: $\forall x_0 \lim_{\substack{x \to x_0 \\ x > x_0}} F(x) = F(x_0)$

Identically Distributed RVs

- Two random variables X_1 and X_2 are identically distributed iif for all sets of values A

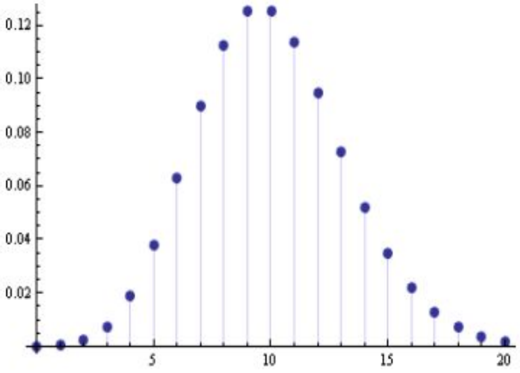
$$P(X_1 \in A) = P(X_2 \in A)$$

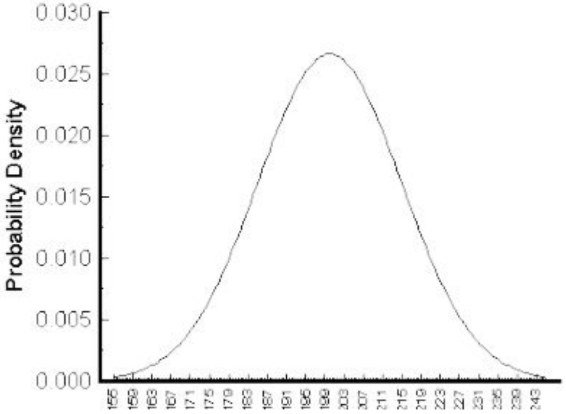
- So that means the variables are equal?
  - NO.
  - Example: Let's toss a coin 3 times and let X_H and X_F represent the number of heads/tails respectively
  - They have the same distribution but $X_H = 1 - X_F$

Discrete vs. Continuous RVs

- Step CDF
- $\mathcal{X}$ is discrete
- Probability mass
  - $f_X(x) = P(X = x) \ \forall x$

- Continuous CDF
- $\mathcal{X}$ is continuous
- Probability density
  - $F_X(x) = \int_{-\infty}^{x} f_X(t)dt \ \ \forall x$

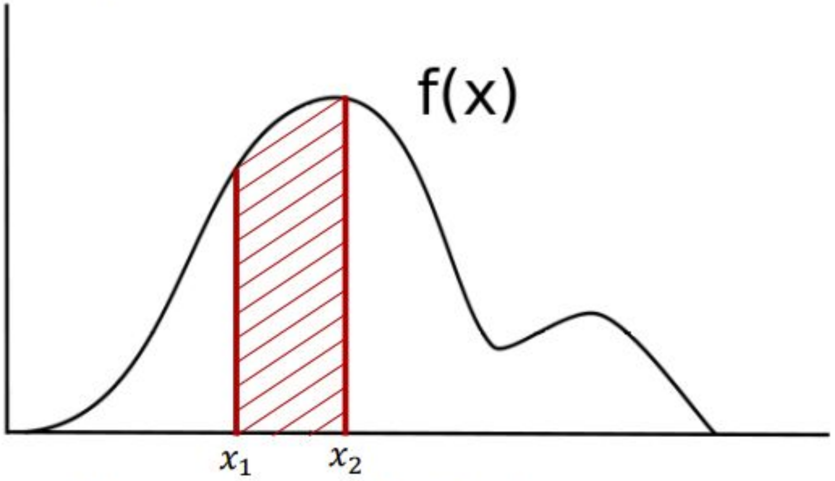




Interval Probabilities

- Obtained by integrating the area under the curve

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f_x(x)dx$$



- This explains why P(X=x) = 0 for continuous distributions!

$$P(X = x) \leq \lim_{\substack{\epsilon \to 0 \\ \epsilon > 0}} [F_x(x) - F_x(x - \epsilon)] = 0$$

Example problem 4

Suppose we toss a fair coin until we get exactly two heads. Describe the sample space S. Calculate the probability mass function of the random variable describing the number of tosses, i.e., calculate the probability that exactly k tosses are required for each possible value of k.

Let A_i be random variable describing the outcome of ith toss and N be random variable describing the total number of tosses. Then,

$$A_i = \begin{cases} 1, \text{if we got head in } i\text{th toss} \\ 0, \text{if we got tail in } i\text{th toss} \end{cases}$$

$$P(A_i = 1) = P(A_i = 0) = \frac{1}{2}$$

The sample space S will be

$$S = \left\{ (A_1, A_2, \ldots, A_N) : \sum_{i=1}^{N-1} A_i = 1, A_N = 1, N \geq 2 \right\}$$

and the probability that exactly k tosses are required will be

$$\begin{aligned} P(N = k) &= \begin{cases} \binom{k-1}{1} P(A_i = 1)^2 P(A_i = 0)^{k-2}, & k = 2, 3, \cdots \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{(k-1)!}{1!(k-2)!} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{k-2}, & k = 2, 3, \cdots \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{k-1}{2^k}, & k = 2, 3, \cdots \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

Expectation

We may consider the **expectation** (or "mean") of a distribution:

$$E(X) = \begin{cases} \sum_x x f_X(x) & \text{X is discrete} \\ \int_{-\infty}^{\infty} x f_X(x)\, dx & \text{X is continuous} \end{cases}$$

Expectation is linear:

$$\begin{aligned} E(aX + bY + c) &= \sum_{x,y} (ax + by + c) f_{X,Y}(x,y) \\ &= \sum_{x,y} ax f_{X,Y}(x,y) + \sum_{x,y} by f_{X,Y}(x,y) + \sum_{x,y} c f_{X,Y}(x,y) \\ &= a \sum_{x,y} x f_{X,Y}(x,y) + b \sum_{x,y} y f_{X,Y}(x,y) + c \sum_{x,y} f_{X,Y}(x,y) \\ &= a \sum_x x \sum_y f_{X,Y}(x,y) + b \sum_y y \sum_x f_{X,Y}(x,y) + c \\ &= a \sum_x x f_X(x) + b \sum_y y f_Y(y) + c \\ &= aE(X) + bE(Y) + c \end{aligned}$$

Variance

We may consider the **variance** of a distribution:

$$\text{Var}(X) = E(X - EX)^2$$

This may give an idea of how "spread out" a distribution is.

A useful alternate form is:

$$\begin{aligned} E(X - EX)^2 &= E[X^2 - 2XE(X) + (EX)^2] \\ &= E(X^2) - 2E(X)E(X) + (EX)^2 \\ &= E(X^2) - (EX)^2 \end{aligned}$$

Variance

Variance is non-linear but the following holds:

$$\text{Var}(aX) = E(aX - E(aX))^2 = E(aX - aEX)^2 = a^2 E(X - EX)^2 = a^2\text{Var}(X)$$

$$\text{Var}(X+c) = E(X+c-E(X+c))^2 = E(X-EX+c-c)^2 = E(X-EX)^2 = \text{Var}(X)$$

$$\begin{aligned}\text{Var}(X+Y) &= E(X - EX + Y - EY)^2 \\ &= \underbrace{E(X - EX)^2}_{\text{Var}(X)} + \underbrace{E(Y - EY)^2}_{\text{Var}(Y)} + 2\underbrace{E(X - EX)(Y - EY)}_{\text{Cov}(X,Y)}\end{aligned}$$

So when X, Y are independent we have:

why?

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$$

Moments (characteristic of a distribution)

- Expectations
  - The expected value of a function g depending on a r.v. $X \sim P$ is defined as $Eg(X) = \int g(x)P(x)dx$
- n^{th} moment of a probability distribution

$$\mu_n = \int x^n P(x)dx$$

- mean $\mu = \mu_1$
- n^{th} central moment

$$\mu_n{}' = \int (x - \mu)^n P(x)dx$$

- Variance $\sigma^2 = \mu_2{}'$

Table of Common Distributions

taken from *Statistical Inference* by Casella and Berger

Discrete Distrbutions

distribution	pmf	mean	variance	mgf/moment
Bernoulli(p)	$p^x(1-p)^{1-x};\ x=0,1;\ p\in(0,1)$	p	$p(1-p)$	$(1-p)+pe^t$
Beta-binomial(n,α,β)	$\binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(\alpha+\beta+n)}$	$\frac{n\alpha}{\alpha+\beta}$	$\frac{n\alpha\beta}{(\alpha+\beta)^2}$	
Notes: If $X\|P$ is binomial (n,P) and P is beta(α,β), then X is beta-binomial(n,α,β).				
Binomial(n,p)	$\binom{n}{x}p^x(1-p)^{n-x};\ x=1,\ldots,n$	np	$np(1-p)$	$[(1-p)+pe^t]^n$
Discrete Uniform(N)	$\frac{1}{N};\ x=1,\ldots,N$	$\frac{N+1}{2}$	$\frac{(N+1)(N-1)}{12}$	$\frac{1}{N}\sum_{i=1}^{N}e^{it}$
Geometric(p)	$p(1-p)^{x-1};\ p\in(0,1)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$
Note: $Y=X-1$ is negative binomial$(1,p)$. The distribution is *memoryless*: $P(X>s\|X>t)=P(X>s-t)$.				
Hypergeometric(N,M,K)	$\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}};\ x=1,\ldots,K$	$\frac{KM}{N}$	$\frac{KM}{N}\frac{(N-M)(N-k)}{N(N-1)}$	?
	$M-(N-K)\le x\le M;\ N,M,K>0$			
Negative Binomial(r,p)	$\binom{r+x-1}{x}p^r(1-p)^x;\ p\in(0,1)$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$	$\left(\frac{p}{1-(1-p)e^t}\right)^r$
	$\binom{y-1}{r-1}p^r(1-p)^{y-r};\ Y=X+r$			
Poisson(λ)	$\frac{e^{-\lambda}\lambda^x}{x!};\ \lambda\ge 0$	λ	λ	$e^{\lambda(e^t-1)}$
Notes: If Y is gamma(α,β), X is Poisson$(\frac{x}{\beta})$, and α is an integer, then $P(X\ge\alpha)=P(Y\le y)$.				

Continuous Distributions

distribution	pdf	mean	variance	mgf/moment
Beta(α,β)	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1};\ x\in(0,1),\ \alpha,\beta>0$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$1+\sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}\right)\frac{t^k}{k!}$
Cauchy(θ,σ)	$\frac{1}{\pi\sigma}\frac{1}{1+(\frac{x-\theta}{\sigma})^2};\ \sigma>0$	does not exist	does not exist	does not exist
Notes: Special case of Students's t with 1 degree of freedom. Also, if X,Y are iid $N(0,1)$, $\frac{X}{Y}$ is Cauchy				
χ^2_p	$\frac{1}{\Gamma(\frac{p}{2})2^{\frac{p}{2}}}x^{\frac{p}{2}-1}e^{-\frac{x}{2}};\ x>0,\ p\in N$	p	$2p$	$\left(\frac{1}{1-2t}\right)^{\frac{p}{2}},\ t<\frac{1}{2}$
Notes: Gamma$(\frac{p}{2},2)$.				
Double Exponential(μ,σ)	$\frac{1}{2\sigma}e^{-\frac{\lvert x-\mu\rvert}{\sigma}};\ \sigma>0$	μ	$2\sigma^2$	$\frac{e^{\mu t}}{1-(\sigma t)^2}$
Exponential(θ)	$\frac{1}{\theta}e^{-\frac{x}{\theta}};\ x\geq 0,\ \theta>0$	θ	θ^2	$\frac{1}{1-\theta t},\ t<\frac{1}{\theta}$
Notes: Gamma$(1,\theta)$. Memoryless. $Y=X^{\frac{1}{\gamma}}$ is *Weibull*. $Y=\sqrt{\frac{2X}{\beta}}$ is *Rayleigh*. $Y=\alpha-\gamma\log\frac{X}{\beta}$ is *Gumbel*.				
F_{ν_1,ν_2}	$\frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}\frac{x^{\frac{\nu_1-2}{2}}}{\left(1+(\frac{\nu_1}{\nu_2})x\right)^{\frac{\nu_1+\nu_2}{2}}};\ x>0$	$\frac{\nu_2}{\nu_2-2},\ \nu_2>2$	$2(\frac{\nu_2}{\nu_2-2})^2\frac{\nu_1+\nu_2-2}{\nu_1(\nu_2-4)},\ \nu_2>4$	$EX^n=\frac{\Gamma(\frac{\nu_1+2n}{2})\Gamma(\frac{\nu_2-2n}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})}\left(\frac{\nu_2}{\nu_1}\right)^n,\ n<\frac{\nu_2}{2}$
Notes: $F_{\nu_1,\nu_2}=\frac{\chi^2_{\nu_1}/\nu_1}{\chi^2_{\nu_2}/\nu_2}$, where the χ^2s are independent. $F_{1,\nu}=t^2_\nu$.				
Gamma(α,β)	$\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-\frac{x}{\beta}};\ x>0,\ \alpha,\beta>0$	$\alpha\beta$	$\alpha\beta^2$	$\left(\frac{1}{1-\beta t}\right)^\alpha,\ t<\frac{1}{\beta}$
Notes: Some special cases are exponential ($\alpha=1$) and χ^2 ($\alpha=\frac{p}{2},\beta=2$). If $\alpha=\frac{2}{3}$, $Y=\sqrt{\frac{X}{\beta}}$ is *Maxwell*. $Y=\frac{1}{X}$ is *inverted gamma*.				
Logistic(μ,β)	$\frac{1}{\beta}\frac{e^{-\frac{x-\mu}{\beta}}}{\left[1+e^{-\frac{x-\mu}{\beta}}\right]^2};\ \beta>0$	μ	$\frac{\pi^2\beta^2}{3}$	$e^{\mu t}\Gamma(1+\beta t),\ \lvert t\rvert<\frac{1}{\beta}$
Notes: The cdf is $F(x\mid\mu,\beta)=\frac{1}{1+e^{-\frac{x-\mu}{\beta}}}$.				
Lognormal(μ,σ^2)	$\frac{1}{\sqrt{2\pi}\sigma}\frac{1}{x}e^{-\frac{(\log x-\mu)^2}{2\sigma^2}};\ x>0,\sigma>0$	$e^{\mu+\frac{\sigma^2}{2}}$	$e^{2(\mu+\sigma^2)}-e^{2\mu+\sigma^2}$	$EX^n=e^{n\mu+\frac{n^2\sigma^2}{2}}$
Normal(μ,σ^2)	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}};\ \sigma>0$	μ	σ^2	$e^{\mu t+\frac{\sigma^2t^2}{2}}$
Pareto(α,β)	$\frac{\beta\alpha^\beta}{x^{\beta+1}};\ x>\alpha,\ \alpha,\beta>0$	$\frac{\beta\alpha}{\beta-1},\ \beta>1$	$\frac{\beta\alpha^2}{(\beta-1)^2(\beta-2)},\ \beta>2$	does not exist
t_ν	$\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\frac{1}{\sqrt{\nu\pi}}\frac{1}{(1+\frac{x^2}{\nu})^{\frac{\nu+1}{2}}}$	$0,\ \nu>1$	$\frac{\nu}{\nu-2},\nu>2$	$EX^n=\frac{\Gamma(\frac{\nu+1}{2})\Gamma(\nu-\frac{n}{2})}{\sqrt{\pi}\Gamma(\frac{\nu}{2})}\nu^{\frac{n}{2}},\ n$ even
Notes: $t^2_\nu=F_{1,\nu}$.				
Uniform(a,b)	$\frac{1}{b-a},\ a\leq x\leq b$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt}-e^{at}}{(b-a)t}$

Example problem 5

Let $X_1, X_2, \ldots, X_n$ be random variables all with unknown distributions, whose mean is 0, their deviation is 1 and all are independent from one another. Let $Y = \frac{X_1 + X_2 + \ldots + X_n}{n}$. Find $E[Y]$ and $\sigma(Y)$.

Solution

Y can also be written $\frac{\sum X}{n}$.

The expectation of Y is $E(\frac{\sum X}{n}) = \frac{1}{n}E(\sum X) = \frac{1}{n}\sum E(X) = \frac{1}{n}0 = 0$, in accordance with your intuition.

The variance of Y is $var(Y) = var(\frac{\sum X}{n}) = \frac{1}{n^2}var(\sum X) = \frac{1}{n^2}\sum var(X) = \frac{n}{n^2} = \frac{1}{n}$. Standard deviation is the square root of this, $\sigma(Y) = \frac{1}{\sqrt{n}}$.

Example problem 6

Suppose we generate a random variable X in the following way. First we flip a fair coin. If the coin is heads, take X to have a U(0,1) distribution. If the coin is tails, take X to have a U(3,4) distribution.

(a) Find the mean of X.

(b) Find the standard deviation of X.

Solution

(a) Let $\epsilon \sim \text{Bern}(0.5)$, $U_1 \sim U(0,1)$ with $\mathbb{E}[U_1] = \frac{1}{2}$ and $\text{Var}[U_1] = \frac{1}{12}$, $U_2 \sim U(3,4)$ with $\mathbb{E}[U_2] = \frac{7}{2}$ and $\text{Var}[U_2] = \frac{1}{12}$. Then $X = \mathbb{1}(\epsilon = 1)U_1 + \mathbb{1}(\epsilon = 0)U_2$.

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}_\epsilon\left[\mathbb{E}\left[U \mid \epsilon\right]\right] \\
&= p(\epsilon = 1)\mathbb{E}[U \mid \epsilon = 1] + p(\epsilon = 0)\mathbb{E}[U \mid \epsilon = 0] \\
&= 0.5\mathbb{E}[U_1] + 0.5\mathbb{E}[U_2] \\
&= 2
\end{aligned}$$

(b) We have $\mathbb{E}[U_1^2] = \text{Var}[U_1] + \mathbb{E}[U_1]^2 = \frac{1}{3}$ and $\mathbb{E}[U_2^2] = \text{Var}[U_2] + \mathbb{E}[U_2]^2 = \frac{37}{3}$.

$$\begin{aligned}
\text{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \mathbb{E}_\epsilon[\mathbb{E}[U^2 \mid \epsilon]] - \mathbb{E}[X]^2 \\
&= 0.5\mathbb{E}[U_1^2] + 0.5\mathbb{E}[U_2^2] - 4 \\
&= \frac{7}{3}
\end{aligned}$$

Thus, the standard deviation of X is $\sqrt{\frac{7}{3}} \approx 1.5275$.

Practice problem (source: 18751)

To the best of our knowledge, with probability 0.8 Al is guilty of the crime for which he is about to be tried. Bo and Ci, each of whom knows whether or not Al is guilty, have been called to testify. Bo is a friend of Al s and will tell the truth if Al is innocent but will lie with probability 0.2 if Al is guilty. Ci hates everybody but the judge and will tell the truth if Al is guilty but will lie with probability 0.3 if Al is innocent.

Given this model of the physical situation:

1. Determine the probability that the witnesses give conflicting testimony.
2. Determine the probability that Bo commits perjury, determine the probability that Ci commits perjury.
3. What is the conditional probability that Al is innocent given that Bo and Ci gave conflicting testimony?
4. Are the events {Bo tells a lie} and {Ci tells a lie} independent? Are these events conditionally independent to an observer who knows whether or not Al is guilty?