Introduction to Probability
Approximations of the binomial distribution

Bin(n, p) with large n

The PDF of the discrete binomial random variable $Bin(n, p)$,

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k},\ \ k = 0, 1, 2, ..., n-1, n,$$

shows one of two very particular patterns in its shape as the number of independent trials n gets larger.

Bin(n, p) with large n: approximations

If p is close to 0.5, this shape appears like a *bell curve* with exponentially-falling tails, much like a **Gaussian** distribution.

If p is close to 0, this shape appears like a **Poisson** distribution.

We will see details on how well we can use these distributions to approximate binomials with large n.

Approximations of RVs by easier-to-calculate RVs

Recall some properties of $S_n \sim Bin(n, p)$:

- $E(S_n) = np$
- $Var(S_n) = np(1-p)$
- $S_n \sim \sum_{i=1}^{n} X_i$, where $X_i \sim Bern(p)$ are IID.

We will examine, for n large:

- $Bin(n, p)$ approximated by a normal (when p is near $\frac{1}{2}$), via the **de Moivre-Laplace Theorem** (a special case of the **Central Limit Theorem**), and
- $Bin(n, p)$ approximated by a Poisson (when p is near 0), via the **law of rare events**.

Normal (Gaussian) random variable: PDF

Normal (Gaussian) random variables are used to model vast amounts of physical, social, and financial processes.

$X \sim N(\mu, \sigma^2)$ (with $\mu = E(X)$ and $\sigma^2 = Var(X)$) if X has PDF

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

$Z \sim N(0, 1)$ is called a **standard normal random variable**.

Normalizing constant of the normal PDF

Since any PDF f must satisfy

$$\int_{-\infty}^{\infty} f(x)dx = 1,$$

then we have the following from the standard normal PDF:

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}.$$

Normal (Gaussian) random variable: CDF

The CDF of $X \sim N(\mu, \sigma^2)$,

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt,$$

has no closed form (no simple formula). As this is a very important distribution, we need a way to calculate this CDF in a reasonable fashion.

The CDF of $Z \sim N(0,1)$ also has no closed form, but is denoted $N(x)$ or $\Phi(x)$. It is defined by

$$N(x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

Calculations for $\Phi(x)$ are listed on a **z-table**.

Standardization

A **standard** random variable has mean 0 and variance 1.

For a random variable X with mean $E(X) = \mu$ and variance $Var(X) = \sigma^2$, the **standardized** version of X is

$$Z = \frac{X - \mu}{\sigma}.$$

We use this transformation often in discussing normal random variables, to turn questions about probabilities for X into probabilities for Z (which can then be read off the Z-table).

de Moivre-Laplace (Central Limit) Theorem

One of the fundamental limit theorems of probability theory is the **Central Limit Theorem** (CLT); when stated for binomial RVs, this is called the **de Moivre-Laplace Theorem**:

de Moivre-Laplace (Central Limit) Theorem

Theorem

If $S_n \sim Bin(n,p)$ *with fixed* $0 < p < 1$. *Then, with*

$$\mu_n = E(S_n) = np \textit{ and } \sigma_n^2 = Var(S_n) = np(1-p),$$

and the standardized RVs

$$Z_n = \frac{S_n - \mu_n}{\sigma_n},$$

we have convergence in distribution to a standard normal random variable: $Z_n \to Z \sim N(0,1)$ *as* $n \to \infty$*; for any* $-\infty \le a \le b \le \infty$,

$$\lim_{n\to\infty} F_{Z_n}(b) - F_{Z_n}(a) = \lim_{n\to\infty} P\left(a < Z_n \le b\right) = \Phi(b) - \Phi(a).$$

Binomial Approximation via normal distribution

CLT approximation is precisely the motivation behind many important models in finance (for example, the **binomial tree** model for stock price evolution with a large number of coin flips n).

As a rule of thumb, we'll expect the approximation

$$P\left(a < \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \approx \Phi(b) - \Phi(a)$$

of the standardization of $S_n \sim Bin(n, p)$ to be good if

$$Var(S_n) = np(1-p) > 10.$$

Binomial Approximation via normal distribution

Example

Let's play a game.

I flip a fair coin 100 times: if it lands heads 60 or more times, you win \$10. If not, you win nothing.

What is a fair price for this game?

(This is a simplified model for a special type of stock option called a **European digital call option**.)

Binomial Approximation via normal distribution

Let $S_n \sim Bin(n = 100, p = 1/2)$ count the number of heads on flipped.

Since $np(1-p) = 100(0.5)(0.5) = 25 > 10$, we can reasonably approximate these probabilities via CLT. Specifically, we want

$$
\begin{aligned}
P(S_n \geq 60) &= P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \geq \frac{60 - np}{\sqrt{np(1-p)}}\right) \\
&\approx P\left(Z \geq \frac{60-50}{5}\right) \\
&= P(Z \geq 2) = 1 - N(2) \approx 1 - 0.9772 = 0.0228.
\end{aligned}
$$

Applications of the CLT: Binomial Approximation

According to this approximation, the probability of winning \$10 is approximately 0.0228, and so the probability of winning nothing is approximately 1 - 0.0228 $= 0.9772$.

Hence, the fair price of the game is

$$0.0228(10) + 0.9772(0) = 0.228,$$

or about 22.8 cents for a small house advantage.

Applications of the CLT: Binomial Approximation

The actual probability is

$$P(Y \geq 60) = \sum_{j=60}^{100} \binom{100}{j} (0.5)^{100} \approx 0.028444.$$

The difference between 28.44 cents (actual) and 22.8 cents seems high - is there a way to make the approximation better than this?

Continuity Correction for Discrete Random Variables

Continuity correction adjusts the binomial approximation slightly, by accounting for the discrepancy between a smooth PDF curve's integral and the blocky rectangles of a histogram.

To use the continuity correction, add or subtract 0.5 in the appropriate direction to the integer value you are approximating: if $a, b \in \mathbb{Z}$, use the correction on $S_n \sim Bin(n, p)$ to check

$$P(a \leq S_n \leq b) = P(a - 0.5 \leq S_n \leq b + 0.5).$$

We will apply the continuity correction to our example.

Continuity Correction: Digital Option Attempt #2

$P(S_n \geq 60)$ is better approximated by $P(S_n \geq 59.5)$.

$$\begin{aligned} P(S_n \geq 60) &\approx P\left(Z \geq \frac{59.5 - 50}{5}\right) \\ &= P(Z \geq 1.9) = 1 - N(1.9) \approx 1 - 0.9713 = 0.0287, \end{aligned}$$

giving a much closer approximation of a fair price of 28.7 cents against the actual computation of 28.44 cents.

"three σ rule"

Recall the standard deviation (σ) distances from the either side of the mean:

$$\begin{aligned}
\Phi(1) - \Phi(-1) &\approx 2(0.3413) = 0.6826 \quad \text{(1 std dev from mean)} \\
\Phi(2) - \Phi(-2) &\approx 2(0.4772) = 0.9554 \quad \text{(2 std dev from mean)} \\
\Phi(3) - \Phi(-3) &\approx 2(0.4987) = 0.9974 \quad \text{(3 std dev from mean)}
\end{aligned}$$

These numbers give good guidelines for how much of a distribution to expect a certain distance from the mean.

The so-called "$3 - \sigma$ rule" says that the vast majority - over 99.7% - of the values of a normal random variable will land within 3 standard deviations of the mean.

Stirling's Approximation

Use the notation $a_n \sim b_n$ to mean that the sequences of real numbers (a_n) and (b_n) are **asymptotically equal** to each other; that is,

$$a_n \sim b_n \iff \lim_{n\to\infty} \frac{a_n}{b_n} = 1.$$

This notation is used to describe an important result, known as **Stirling's approximation**, or **Stirling's formula**, which relates approximates $n!$ for large n in terms of exponentials:

$$\textbf{Stirling's formula}: \; n! \sim n^n e^{-n}\sqrt{2\pi n}.$$

Sketch of Proof of de Moivre-Laplace Theorem

Stirling's formula can be used to approximate binomial probabilities with exponentials: for large n, we have, with $q = 1 - p$,

$$\begin{aligned} \binom{n}{k} p^k q^{n-k} &= \frac{n! p^k q^{n-k}}{k!(n-k)!} \\ &\sim \frac{n^n e^{-n} \sqrt{2\pi n} p^k q^{n-k}}{k^k e^{-k} \sqrt{2\pi k} (n-k)^{n-k} e^{-(n-k)} \sqrt{2\pi(n-k)}} \\ &\approx \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k-np)^2}{2npq}\right), \end{aligned}$$

which, when applied to our standardized binomial probability, gives

$$P\left(a < \frac{S_n - np}{\sqrt{npq}} \le b\right) = \sum_{k=\lceil np + a\sqrt{npq} \rceil}^{\lfloor np + b\sqrt{npq} \rfloor} \binom{n}{k} p^k q^{n-k} \approx \Phi(b) - \Phi(a).$$

Law of Large Numbers (binomial)

Various **Laws of Large Numbers** (LLN) states that a sum of IID random variables with finite mean have their **sample average** converge to their mean.

We will first see this for binomials:

Theorem

For any $\varepsilon > 0$, *and* $S_n \sim Bin(n, p)$,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) = 1.$$

Law of Large Numbers (binomial)

That is, the probability that the sample average

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j = \frac{S_n}{n}$$

is within ε of the probabilistic average, i.e. mean $\mu = p$ of $X_i \sim Bern(p)$, goes to 1 as the number of samples n increases.

Confidence Intervals

The $100r\%$ **confidence interval** of a random variable X is the interval, centered on its mean, inside which the probability of that variable being in that interval is r.

We can compute confidence intervals to know when we can reliably (up to the percent we have chosen) believe an estimated value is "close" to its "true mean".

Confidence Intervals

For example, we know that, for large n, the sample average of n $Bern(p)$ trials is $\frac{S_n}{n}$, with $S_n \sim Bin(n, p)$. By the LLN,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) = 1.$$

Presume that we do not know p, but we run many trials to estimate p. If we call $\hat{p} = \frac{S_n}{n}$ an **estimator** of p, how many trials must we run to be "reasonably sure" that $\hat{p}$ is "close" to p?

That question can be rigorously answered with a fixed $\varepsilon > 0$, the LLN, and the CLT.

Confidence Intervals

Let's examine

$$P\left(|\hat{p} - p| < \varepsilon\right)$$

for large n. Rewriting and using the CLT, we see

$$\begin{aligned}
P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) &= P\left(-\varepsilon < \frac{S_n}{n} - p < \varepsilon\right) \\
&= P\left(-n\varepsilon < S_n - np < n\varepsilon\right) \\
&= P\left(\frac{-n\varepsilon}{\sqrt{np(1-p)}} < \frac{S_n - np}{\sqrt{np(1-p)}} < \frac{n\varepsilon}{\sqrt{np(1-p)}}\right) \\
&\approx \Phi\left(\frac{n\varepsilon}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{-n\varepsilon}{\sqrt{np(1-p)}}\right) \\
&= 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1.
\end{aligned}$$

Confidence Intervals

The normal approximation to the binomial probability with $\hat{p} = \frac{S_n}{n}$,

$$P\left(|\hat{p} - p| < \varepsilon\right) \approx 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1,$$

can be simplified further.

Since we do not know p, we can get a lower bound for all values of p. First, note that, for $0 < p < 1$, the "most random" a coin flip can be is fair, i.e. $p = \frac{1}{2}$.

This intuition (or, better, some calculus) yields

$$\sqrt{p(1-p)} \leq \sqrt{\frac{1}{4}} = \frac{1}{2} \implies \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \geq 2\varepsilon\sqrt{n}.$$

Confidence Intervals

Thus, since the CDF Φ is increasing, we have a bound that works for all p:

$$P\left(|\hat{p} - p| < \varepsilon\right) \geq 2\Phi\left(2\varepsilon\sqrt{n}\right) - 1.$$

The $100r\%$ **confidence interval** that our probability estimator $\hat{p}$ is within ε of its true probability p is the interval $(\hat{p} - \varepsilon, \hat{p} + \varepsilon)$ such that

$$P\left(|\hat{p} - p| < \varepsilon\right) \geq 2\Phi\left(2\varepsilon\sqrt{n}\right) - 1 \geq r,$$

which reduces to

$$100r\% \text{ confidence that } p \in (\hat{p} - \varepsilon, \hat{p} + \varepsilon): \;\; \Phi\left(2\varepsilon\sqrt{n}\right) = \frac{1 + r}{2}$$

in four parameters: ε, r, $\hat{p}$, and n.

Confidence Intervals

Different questions can be asked about confidence intervals, and all deal with having three parameters of

$$100r\% \text{ confidence that } p \in (\hat{p} - \varepsilon, \hat{p} + \varepsilon) : \ \Phi\left(2\varepsilon\sqrt{n}\right) = \frac{1+r}{2}$$

and solving for the fourth.

Confidence Intervals

$$100r\% \text{ confidence that } p \in (\hat{p}-\varepsilon, \hat{p}+\varepsilon): \; \Phi\left(2\varepsilon\sqrt{n}\right) = \frac{1+r}{2}$$

- How many trials n must be run to be $100r\%$ confident our sample probability $\hat{p}$ is within ε of the true probability p? (Compute n.)
- Find the $100r\%$ confidence interval for the true probability p after n trials yields sample probability $\hat{p}$. (Compute ε).
- How confident are we that p is in the given confidence interval? (Compute r.)

Example: How Many Die Rolls to Be Reasonably Certain?

We believe we have a fair die (with "true probability" $p = \frac{1}{6}$ of rolling a 2).

How many times should we roll the die to be at least 95% confident that the relative frequency of a 2 appearing is within 0.01 of its actual probability of $\frac{1}{6}$?

Example: How Many Die Rolls to Be Reasonably Certain?

Let's pose the question with the variables we know:

How many times n should we roll the die to be at least $100r = 95\%$ certain that the relative frequency $\hat{p}$ of a 2 appearing is within $\varepsilon = 0.01$ of its actual probability of $p = \frac{1}{6}$?

Example: How Many Die Rolls to Be Reasonably Certain?

$$100r\% \text{ confidence that } p \in (\hat{p} - \varepsilon, \hat{p} + \varepsilon) : \ \Phi\left(2\varepsilon\sqrt{n}\right) = \frac{1+r}{2}$$

becomes

$$95\% \text{ confidence that } \frac{1}{6} \in (\hat{p} - 0.01, \hat{p} + 0.01)$$

with

$$\Phi\left(0.02\sqrt{n}\right) \geq \frac{1.95}{2} = 0.975 \implies 0.02\sqrt{n} \approx 1.96,$$

which gives $\sqrt{n} \geq \frac{1.96}{0.02} = 98$, or $n \geq 98^2 = 9604$ rolls.

Example: How Many Die Rolls to Be Reasonably Certain?

We can improve this estimate if we don't use the worst case scenario for $\sqrt{p(1-p)} \leq \frac{1}{2}$, and instead plug in $p = \frac{1}{6}$. This yields

$$95\% \text{ confidence that } \frac{1}{6} \in (\hat{p} - 0.01, \hat{p} + 0.01)$$

with

$$\Phi\left(\frac{0.01\sqrt{n}}{\sqrt{\frac{1}{6}\cdot\frac{5}{6}}}\right) = 0.975$$

$$\implies \Phi\left(\frac{0.06\sqrt{n}}{\sqrt{5}}\right) \geq 0.975 \implies \frac{0.06\sqrt{n}}{\sqrt{5}} \geq 1.96,$$

which gives the much better estimate $\sqrt{n} \geq \frac{1.96\sqrt{5}}{0.06} \approx 73.044887$, or $n \geq 5336$ rolls.

Maximum Likelihood Estimation

If we take p to be a variable parameter, we can do calculus on a PMF or PDF that uses p to determine the most "likely" value for a random variable to take.

For a random variable X whose PMF or PDF uses a parameter we will call p, define the **likelihood function**

$$L(p) = P(X = k)$$

as a function of $p \in [0, 1]$, for *fixed* k.

Maximum Likelihood Estimation

Then calculus derivative tests can tell us the maximum of $L(p)$, which is the p that makes k the most likely value X will take.

We call the value $\hat{p} = \sup_{p \in [0,1]}$ the **maximum likelihood estimator (MLE)**, and it should be clear that $\hat{p}$ is a function of k.

Maximum Likelihood Estimation

In our earlier die-rolling example, we had the sample average $\hat{p} = \frac{S_n}{n}$ as our random variable, with $S_n \sim Bin(n, p)$.
For a fixed k, the likelihood function is

$$L(p) = P\left(\frac{S_n}{n} = \frac{k}{n}\right) = \binom{n}{k} p^k (1-p)^{n-k},$$

whose derivative with respect to p is at a critical point if

$$\begin{aligned} L'(p) &= \binom{n}{k}\left(kp^{k-1}(1-p)^{n-k} - (n-k)(1-p)^{n-k-1}p^k\right) \\ &= \binom{n}{k} p^{k-1}(1-p)^{n-k-1}(k - np) = 0. \end{aligned}$$

The first derivative test says $L(p)$ has a maximum at $\hat{p} = \frac{k}{n} = \frac{S_n}{n}$, which (again) justifies its use as an estimator for p.

Example: polling

Random **polling** is, ideally, modeled like drawing balls from urns *without replacement* (since you must not ask the same person twice).

However, pulling people from a very large population offers a very small chance we happen to double up, and so we will maintain our near-independence assumption.

Example: polling

Say we poll $n = 1061$ people*, at random from the population†, asking if they believe the US moon landing in 1969 was real or staged. 64 of them say it was staged.

What is the 95% confidence interval, using our sample average of $\hat{p} = \frac{64}{1061} \approx 6\%$ in this poll, that captures the overall population's "true belief probability" p on this question?

---

*http://news.gallup.com/poll/3712/landing-man-moon-publics-view.aspx, July 20, 1999

†No poll is "perfect". Poll design and execution is exceedingly difficult.

Confidence in the moon landing

We assume that $S_n \sim Bin(n, p)$ is the number of people in our poll that believe the moon landing was staged, and set our sample average as

$$\hat{p} = \frac{S_n}{n} = \frac{64}{1061} \approx 0.06.$$

Confidence in the moon landing

We want to compute ε such that, for "true" belief percentage p,

$$\text{we have 95\% confidence that } p \in (0.06 - \varepsilon, 0.06 + \varepsilon)$$

with worst case scenario

$$\Phi\left(2\varepsilon\sqrt{n}\right) \geq 0.975 \implies \varepsilon \geq \frac{1.96}{2\sqrt{1061}} \approx 0.03.$$

This gives a 95% confidence interval of $p \in (0.03, 0.09)$.[‡]

[‡]Compare: `http://www.rasmussenreports.com/public_content/lifestyle/general_lifestyle/july_2014/have_we_got_a_conspiracy_for_you_9_11_jfk_obama_s_citizenship`, which claim a 2012 poll with $n = 1000$ and $p \in (0.02, 0.08)$, and a 2014 poll with $p \in (0.11, 0.17)$, both at 95% confidence. Polling is hard.

Poisson random variables

X is called a **Poisson random variable** with parameter λ (written Poisson(λ)) if the PMF of X is

$$p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!},\ k = 0, 1, 2, 3, ...$$

This counts the number of "rare events" with "average" rate of occurrence λ.

We can use a Poisson(λ) to approximate a Bin(n, p), using $\lambda = np$, if the number of trials n is "large" and success probability p is "small".

Properties of Poisson random variables

For $X \sim Poisson(\lambda)$,

- $E(X) = \lambda$
- $Var(X) = \lambda$
- $P(X = 0) = e^{-\lambda}$
- $P(X > 0) = 1 - e^{-\lambda}$
- **Poisson approximation of binomial**: If $X \sim Bin(n, p)$ and $Y \sim Poisson(\lambda = np)$, then for any subset $A \subseteq \{0, 1, 2, ...\}$,

$$|P(X \in A) - P(Y \in A)| \leq np^2.$$

Thus, if p is "small" relative to n, Y can be considered a good approximation to X.

Binomial approximated by Poisson: "law of rare events"

If the number of IID "success/failure" trials n is large, and the "success" probability p is small, then we can approximate extreme events of a $Bin(n, p)$ RV by a Poisson($\lambda = np$).

Example

Assume the probability of a nuclear reactor accident each year is $p = 10^{-5}$ ("US utility requirements are 1 in 100,000 years"), there are 100 reactors in operation in the US, and each reactor operates independently each year.[§]

Approximate the probability that there is an accident some time in the next 25 years.

[§] Basic data from http://www.world-nuclear.org/information-library/safety-and-security/safety-of-plants/safety-of-nuclear-power-reactors.aspx. Independent operation and 100 instead of 99 plants are simplifying assumptions.

Binomial approximated by Poisson: "law of rare events"

Let X = the number of accidents.

The actual probability we want, using these numbers, is for $n = 100(25) = 2500$ "trials" of "waiting a year on each reactor for an accident", with "success" probability $p = 10^{-5}$. Then,

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{2500}{0}(1 - 10^{-5})^{2500}$$
$$= 1 - 0.99999^{2500} \approx 1 - 0.97531 = 0.02469,$$

just under 2.5%.

Binomial approximated by Poisson: "law of rare events"

This actual probability might suffer from minor roundoff error, depending on the algorithm used to calculate it.

Approximating via $Y \sim \text{Poisson}(\lambda = np = \frac{2500}{10^5} = 0.025)$,

$$P(X \geq 1) \approx P(Y > 0) = 1 - P(Y = 0) = 1 - e^{-\lambda} \approx 0.02469.$$

Binomial approximated by Poisson: "law of rare events"

If, instead, we use the sample average[¶]

$$\hat{p} = \frac{2}{18000} \approx 0.00011$$

instead of $p = 10^{-5}$, our estimates increase sharply:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{2500}{0}(1 - 0.00011)^{2500}$$
$$= 1 - 0.999889^{2500} \approx 0.242547.$$

[¶](generously referring to only 2 of the 3 major accidents that have occured in "over 17,000 cumulative reactor-years of commercial nuclear power operation in 33 countries")

Binomial approximated by Poisson: "law of rare events"

The Poisson approximation is, with $\lambda = 2500\hat{p} \approx 0.2777778$,

$$P(Y > 0) = 1 - P(Y = 0) = 1 - e^{-\lambda} \approx 0.242535.$$

Note that

$$n\hat{p}^2 \approx 2500(0.00011)^2 \approx 0.000031,$$

so the Poisson approximation is (again) validated.

Geometric random variables

X is called a **geometric random variable** *with parameter* p (written $\mathrm{Geom}(p)$) if its PMF is

$$p_X(k) = p(1-p)^{k-1},\ k = 1, 2, \ldots.$$

This RV represents the number of trials up to a "success" in a run of repeated IID experiments with "success" probability p. That is, $k-1$ "failures", and then "success" on trial k.

Example

"Roll a 5 on a die" has success probability $\frac{1}{6}$, and so failure probability $\frac{5}{6}$. Thus, for $X \sim \mathrm{Geom}(\frac{1}{6})$, the probability it takes exactly 4 rolls to get the first 5 is

$$p_X(4) = \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^3.$$

Exponential random variables

An **exponential** random variable models the amount of time it takes for a given impending event to occur as exponential decay. $X \sim Exp(\lambda)$ has PDF

$$f(x) = \lambda e^{-\lambda x} 1_{(0,\infty)}(x).$$

If $X \sim Exp(\lambda)$, then its CDF is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} \lambda e^{-\lambda t} 1_{(0,\infty)}(t) dt = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x > 0. \end{cases}$$

Properties of geometric and exponential random variables

If $X \sim Geom(p)$ and $T \sim Exp(\lambda)$, then

- $E(X) = \frac{1}{p}$
- $E(T) = \frac{1}{\lambda}$
- $Var(X) = \frac{1-p}{p^2}$
- $Var(T) = \frac{1}{\lambda^2}$

Properties of geometric and exponential random variables

- If T_n is a RV such that $nT_n \sim Geom(\frac{\lambda}{n})$ for large enough n, then
$$\lim_{n\to\infty} P(T_n > t) = e^{-\lambda t},$$
i.e. $T_n \to T$ in distribution ("weakly").
- Geometric random variables are the only discrete random variables that have a special property called **memorylessness**.
- Exponential random variables are the only continuous random variables that have the **memorylessness** property.

Memorylessness of Exponential Random Variables

Memorylessness: If $X \sim Exp(\lambda)$ (a stopwatch for some event), then for $s, t \geq 0$,

$$P(X > s + t \mid X > t) = P(X > s).$$

Informally, the amount of time you've waited for X's event to occur "resets", probabilistically, if you acknowledge that you've waited a certain amount of time.

Memorylessness of Exponential Random Variables

Memorylessness: If $X \sim Exp(\lambda)$, then for $s, t > 0$,

$$P(X > s + t \,|\, X > t) = P(X > s).$$

Proof First, note that $X > s + t \implies X > t$, so

$$P(X > s + t \,|\, X > t) = \frac{P(X > s + t,\, X > t)}{P(X > t)} = \frac{P(X > s + t)}{P(X > t)}.$$

Next, the **right tail probability** of an exponential RV is

$$P(X > t) = \int_t^\infty \lambda e^{-\lambda x} dx = e^{-\lambda t}.$$

Hence,

$$\frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = P(X > s + t \,|\, X > t) = P(X > s) = e^{-\lambda s}. \blacksquare$$

Gamma random variables

Exponential random variables are a special case of **Gamma** random variables.

$X \sim \mathit{Gamma}(\alpha, \beta)$ (with α the "rate" parameter and β the "shape" parameter) if its PDF is

$$f(x) = \frac{\alpha(\alpha x)^{\beta-1} e^{-\alpha x}}{\Gamma(\beta)} 1_{(0,\infty)}(x),$$

where the **Gamma function** is $\Gamma(x) = \int_0^\infty t^x e^{-t} dt$, which is the general form of factorial: $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$.

$\mathit{Exp}(\lambda) \sim \mathit{Gamma}(\lambda, 1)$, and for $n \in \mathbb{N}$, a $\mathit{Gamma}(\lambda, n)$ RV models the sum of n IID $\mathit{Exp}(\lambda)$ (counts the total time for n consecutive events of the same type to occur).