Class Notes 3-6-2019

- **Law of Rare Events or Law of Small Numbers**
  Let $\Lambda > 0$ and consider positive integer n with $\lambda/n < 1$. Let $S_n \sim \text{Bin}(n, \lambda/n)$. Then

$$\lim_{n\to\infty} P(S_n = k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{for } k = 0, 1, 2, \ldots$$

Proof:

$$\begin{aligned} P(S_n = k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{n^k}\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k} \\ &\longrightarrow \frac{\lambda^k}{k!}e^{-\lambda} \qquad \text{as } n\to\infty \text{ for any fixed } k \qquad \Box \end{aligned}$$

Actually a much stronger result giving an upper bound on the approximation error in a much more general setting. To this end we introduce the Poisson-Binomial distribution. Let $X_i \sim \text{Ber}(p_i)$ for $i = 1, 2, \ldots, n$ be independent Bernoulli random variables and denote by $V = X_1 + \ldots + X_n$ the count of successes (ones). This random variable V is said to have a Poisson-Binomial distribution with parameters $p_1, \ldots, p_n$ and it is extremely complicated in its exact form. However, we can approximate it with the distribution of a Poisson random variable $Y \sim \text{Poisson}(\lambda = p_1 + \ldots + p_n)$ and we have the following error bound for the approximation

$$|P(V \in B) - P(Y \in B)| \le 2\sum_{i=1}^n p_i^2 \qquad \text{for any subset } B \subset \{0, 1, 2, 3, \ldots\}$$

An elementary and somewhat lengthy proof is presented on the class web page
http://faculty.washington.edu/fscholz/DATAFILES394_2019/POIBIN.pdf
In fact, with more work the factor 2 in the error bound can be replaced by 1 or $\min(1, 1/\lambda)$ where $\lambda = p_1 + \ldots + p_n$. Thus small p_i and large $p_1 + \ldots + p_n$ drive down the approximation error.

Note that we approximate the very complicated distribution of a Poisson-Binomial random variable with mean $E(V) = E(X_1) + E(X_2) + \ldots + E(X_n) = p_1 + p_2 + p_3 + \ldots + p_n = \lambda$ by the distribution of a Poisson random variable with same mean λ.

The Poisson distribution is useful when counting the number of rare incidences in a large number of trials. The rarity may vary from trial to trial.

For example, spare parts for Boeing aircraft need to be managed properly. The need for a spare part occurs only very rarely, a result of high reliability design. Thus we may want to assess the probability of exceeding the current spare part inventory k, i.e., find $P(V > k) \approx P(Y > k)$. We are dealing with a large fleet of aircraft and a small chance of part failure per plane over a given time period.

The number of times an aircraft gets hit by lightning, or the number of times an aircraft has an aborted landing (has to go around and try again). Aborted landings may vary in frequency from airport to airport, adjusting for m and n landings. Maybe air traffic control is a problem.

The number of times an aircraft has an engine shutdown based on a fleet wide average λ. If a particular airline experiences a high number k of such shutdown we may want to assess

$P(X \geq k)$ assuming that the fleet wide λ applies. That might indicate something amiss with the airline.

Another famous early example of applying the Poisson distribution was provided by Ladislaus Bortkiewicz who tallied the number of deaths by a horse kick in the Prussian army from 1875 to 1894. He wrote a book: Das Gesetz der kleinen Zahlen. (The Law of Small Numbers).

The number of times the ISS (International Space Station) gets hit by space debris or meteoroids of given mass and velocity. That was a big factor in the design of the ISS.

Typos or other types of errors or coincidences are a frequently cited application examples.

- **Factory Accidents:**
  Suppose past records indicate that a factory has on average 3 accidents per month. What is the chance that in a given month we experience more than 6 accidents?

  We have $\lambda = 3$ and

  $$P(V > 6) = 1 - P(V \leq 6) = 1 - \sum_{i=0}^{6} P(V = i) = 1 - \sum_{i=0}^{6} e^{-3}\,\frac{3^i}{i!} = 1 - 0.9665 = 0.0335$$

  Suppose you are told that we had more the 6 accidents per month at least 5 times during the last 5 years. What is the chance of that?

  Let S_{60} be the number of times we have more than 6 accidents per month. $S_{60} \sim \text{Bin}(60, .0335)$ and $P(S_{60} \geq 5) = 1 - P(S_{60} \leq 4) = 0.05047$ and via Poisson approximation using $V \sim \text{Pois}(\lambda = 60 \cdot 0.0335 = 2.01)$ we get $P(V \geq 5) = 1 - P(V \leq 4) = 0.05356$.

  Sometimes the average accident rate λ is given to us indirectly, e.g., the proportion of months without accidents is about 0.2. Treating accidents as rare occurrences and viewing the number V of accidents per month as $V \sim \text{Pois}(\lambda)$ we would use $P(V = 0) = e^{-\lambda} \approx 0.2$ we would use $\lambda \approx -\ln 0.2 = 1.6094$ in subsequent probability calculations, like $P(V \geq 5)$.

- **Comment:**
  While software (e.g. R) easily can give you $P(S_{n=60} \geq 5)$ and even for much larger n, the Poisson approximation is still very useful in situations when the trials have unequal success probabilities $p_1, \ldots, p_n$ and one can only hope to make pronouncements on the average probability $\bar{p} = (p_1 + \ldots + p_n)/n$ and then use $\lambda = n\bar{p}$ for our Poisson approximation. For large n the exact distribution of $X_1 + \ldots + X_n$ is out of reach and it is also unreasonable to know much about the specific values of the required $p_1, \ldots, p_n$.

- **The Exponential Distribution:**
  Let $0 < \lambda < \infty$. Then the continuous random variable X is said to have the exponential distribution with parameter λ, and we write $X \sim \text{Exp}(\lambda)$, if its pdf is

  $$f(x) = \lambda e^{-\lambda x} I_{[0,\infty)}(X)$$

  or equivalently if its cdf is

  $$F(x) = 1 - e^{-\lambda x} \qquad \text{for } x \geq 0 \text{ and } F(x) = 0 \text{ for } x < 0$$

  or equivalently if its complementary survivor function $\bar{F}(x) = 1 - F(x)$ is

  $$\bar{F}(x) = e^{-\lambda x} \qquad \text{for } x \geq 0 \text{ and } \bar{F}(x) = 1 \text{ for } x < 0.$$

Its mean and variance are as follows (integration by parts)

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x} dx = -xe^{-\lambda x}\Big|_0^\infty + \frac{1}{\lambda}\int_0^\infty \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

$$E(X^2) = \int_0^\infty x^2\lambda e^{-\lambda x} dx = -x^2 e^{-\lambda x}\Big|_0^\infty + \frac{1}{\lambda}\int_0^\infty 2x\lambda e^{-\lambda x} dx = \frac{2}{\lambda^2} \;\Rightarrow\; \mathrm{var}(X) = \frac{1}{\lambda^2}$$

- **Memoryless Property of the Exponential Distribution:**
  The exponential distribution is often used as an appropriate distribution when describing the random time T to failure of some device, when failures occur due to external stress events and due to aging of the device. The memoryless property of $X \sim \mathrm{Exp}(\lambda)$ refers to the following property

  $$P(X > t + s | X > t) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} = P(X > s) \qquad \text{for all } s > 0, t > 0$$

  which says that a device lives more than another s units given that it lived more than t units is as if the devices is new and lives more than s units.

  Section 4.7 shows that the exponential distribution is the only distribution on $[0, \infty)$ which has this memoryless property.

- **Exponential Example for the International Space Station (ISS):**
  In 20 years the ISS has survived a few meteoroid impacts. Two noteworthy impacts were recorded in 2012 and 2013. Assume the ISS was designed to survive such impacts with probability .95 over a period of 20 years. What is the chance that it will survive the next 10 years.

  Let T denote the time to impact with major damage (denote this as not surviving). Assume that the ISS had its current size since 1998. By design we have $P(T > 20) = .95 = e^{-\lambda 20} \Rightarrow \lambda = -\ln(.95)/20 = 2.5647 \cdot 10^{-3}$. Then $P(T > 20 + 10 | T > 20) = P(P(T > 10) = e^{-10\ln(.95)/20} = \sqrt{.95} = .9747$.
  For the design work on the ISS, particularly pp 217-231 for my part in it, see
  https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19870019148.pdf
  We supposedly saved about 5% of the weight that needed to be put in orbit. After the Columbia Space Shuttle disaster there was a long delay in space shuttle flights and somehow the probability for no penetration was lowered to 90% over 10 years. Cost increases??