Last time: Expectations of functions of r.v.

Variance of $X$: $\text{Var}(X) = \sigma(X)^2 = \sigma^2(X)$.

Today: Meaning of variance

Chebyshev's Inequality

Poisson r.v.

Interpretation of variance (standard deviation):

If $\sigma(X)$ is small $\rightarrow$ $X$ typically takes values close to its expectation $\mathbb{E}X$.

If $\sigma(X)$ is large $\rightarrow$ $X$ has a non-negligible prob. of taking a value far from its mean.
Explanation: \[ \text{Var}(X) = \mathbb{E} \left( (X - \mathbb{E}X)^2 \right) \]

(Distance of \(X\) from its mean) \(^2\)

Chebyshev's inequality:

\[ \mathbb{P}( |X - \mathbb{E}X| \geq t ) \leq \frac{\text{Var}(X)}{t^2} \quad \text{for any } t > 0. \]

Why is this true?

\[ \text{Var}(X) \geq t^2 \cdot \mathbb{P}( |X - \mathbb{E}X| \geq t ) \]

\[ \sum_{a \in V} (a - \mathbb{E}X)^2 \cdot \mathbb{P}(X=a) \geq \sum_{a \in V} (a - \mathbb{E}X)^2 \cdot \mathbb{P}(X=a) \geq t^2 \cdot \mathbb{P}( |X - \mathbb{E}X| \geq t ) \]

\[ \sum_{a \in V \text{ s.t. } |a - \mathbb{E}X| \geq t} \mathbb{P}(X=a) \]
Example: A factory produces on average 1000 products per day, with a standard deviation of 20 products per day. Give an interval such that you can be 90% certain that the daily productions will fall in this interval.

Sol: Let $X$ be the # of products produced on a given day. We do not know the precise p.m.f. of $X$. We only know:

$\mathbb{E} X = 1000$, $\sigma(X) = 20$. 
This is enough information to answer the question.

We want to find $a, b \in \mathbb{R}$ such that

$$\mathbb{P}(X \in (a, b)) \geq 90\%$$

Note that $X \in (1000 - t, 1000 + t)$

$$\iff |X - \mathbb{E}X| < t.$$  

By Chebyshev's inequality:

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) \leq \frac{\text{Var}(X)}{t^2} = \frac{20^2}{t^2} = \frac{400}{t^2}$$
Want  \[ P( |X-\mu| > t) \leq 10\% \]

\[ = 1 - P( |X-\mu| < t) \]

Want  \[ \frac{400}{t^2} \leq 0.1 \]

\[ \iff t^2 \geq 4000 \]

\[ \iff t \geq 63.25 \]

So  \( t = 64 \) suffices.

Gives the interval:  \[ (936, 1064) \]
The Poisson random variable

Motivation: counts the number of occurrences of infinitely many independent events, each having an infinitesimal prob.

* e.g. the number of accidents per year. (an accident can occur at any moment).

The Poisson r.v. has a parameter \( \lambda > 0 \).

\[ X \sim \text{Poisson}(\lambda), \quad X \in \{0, 1, 2, \ldots \}. \]

p.m.f. of \( X \) is

\[ P(X=k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}. \]

\[ k=0, 1, 2, \ldots \]
Poisson Distribution

- Intensity = 1
- Intensity = 4
- Intensity = 10
- Intensity = 15
(x) Check normalization:
\[
\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \cdot e^{-\lambda} = e^\lambda \cdot e^{-\lambda} = 1
\]

(x) Compute expectation:
\[
\mathbb{E}[X] = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda.
\]

(x) Compute variance:
\[
\mathbb{E}[X(X-1)] = \sum_{k=0}^{\infty} k(k-1) \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2.
\]

\[
\Rightarrow \quad \text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda.
\]
Last time: Chebyshev's inequality
          Poisson r.v.

Today: More about Poisson r.v.

Example: You want to win a prize in a lottery.
          You have $100 to spend.
          Can play in one or two lotteries:
          (A) each ticket costs $1, 10% chance to win.
          (B) each ticket costs $2, 20% chance to win.
          Which do you choose?
Sol: In (A), can buy 100 tickets $\rightarrow X \sim \text{Bin}(100, 0.1)$

In (B), can buy 50 tickets $\rightarrow Y \sim \text{Bin}(50, 0.2)$

So $EX = EY = 10.$

Intuitively, lotteries are "equivalent."

$\rightarrow \mathbb{P}(X=0) = (1-0.1)^{100} = 0.9^{100} = \frac{9^{100}}{10^{100}} = \frac{81^{50}}{10^{100}}.$

$\mathbb{P}(Y=0) = (1-0.2)^{50} = 0.8^{50} = \frac{8^{50}}{10^{50}} = \frac{80^{50}}{10^{100}}.$

$\rightarrow$ So the lotteries give nearly the same chances, but (B) is slightly better.

$\rightarrow$ In fact, $\mathbb{P}(X=k) \propto \mathbb{P}(Y=k)$ for any fixed $k.$
Why is this?

\[ P(0 \text{ of 2 wins in } A) = (1-0.1)^2 = 1 - 0.2 + 0.1^2 \]
\[ P(0 \text{ of 2 wins in } B) = 1 - 0.2 = 0.8 \]

\[ P(1 \text{ of 2 wins in } A) = \binom{2}{1} \cdot 0.1 \cdot (1-0.1) \]
\[ = 0.2 - 2 \cdot 0.1^2 = 0.2 - 0.02 \]
\[ P(1 \text{ of 1 wins in } B) = 0.2 \]

\[ P(2 \text{ of 2 wins in } A) = 0.1^2 = 0.01 \]
\[ P(2 \text{ of 1 wins in } B) = 0 \]
More generally, numerics suggest that the p.m.f. of \( X \sim \text{Bin}(np) \) and \( X' \sim \text{Bin}(n'p') \) are approximately equal if their means are equal, that is, if \( np = n'p' \).

Moreover, the p.m.f. of \( X \) seems to converge as \( n \to \infty \) if \( np \to \text{const.} \).

Precisely, \( \lim_{n \to \infty} \mathbb{P}(\text{Bin}(n, \frac{\lambda}{n}) = k) \) exists for any \( k \geq 0 \) and \( \lambda > 0 \).

Let's compute this limit.
\[
P(\text{Bin}(n, \frac{\lambda}{n}) = k) = \frac{n!}{k!(n-k)!} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1-\frac{\lambda}{n}\right)^{n-k}
\]

\[
= \frac{n!}{k!(n-k)!} \cdot \lambda^k \cdot \frac{1}{n^k} \cdot \left(1-\frac{\lambda}{n}\right)^n \cdot \left(1-\frac{\lambda}{n}\right)^{-k}
\]

\[
= \frac{\lambda^k}{k!} \cdot (1-\frac{\lambda}{n})^n \cdot \frac{n!}{(n-k)!} \cdot \frac{1}{(n-\lambda)^k} \xrightarrow{n \to \infty} \frac{\lambda^k}{k!} e^{-\lambda}
\]

\[
\lim_{n \to \infty} \left(1+\frac{1}{n}\right)^n = e
\]

\[
\frac{n}{n-\lambda} = \frac{1}{1-\frac{\lambda}{n}} \to 1
\]
So \( \Pr(\text{Bin}(n, \frac{1}{n}) = k) \xrightarrow{n \to \infty} \Pr(\text{Poisson}(\lambda) = k) \)

We say \( \text{Bin}(n, \frac{1}{n}) \) r.v. converges to \( \text{Poisson}(\lambda) \) r.v.

Problem: A physicist has a sample of 1 gram of the radioactive isotope \(^{14}\text{C}\) (carbon-14). The atoms of \(^{14}\text{C}\) decay radioactively and emit an electric signal that the physicist detects. He counts an average of 1.4 decays per hour. What, approximately, is \( \Pr(\text{less than 3 decays in next hour}) \)?
Sol: (*1) the sample of ^{14}C contains an unknown, but large, number of atoms, n.

(*1) each atom has a small (unknown) chance p of decaying in the next hour.

(*1) it is reasonable to assume that atoms decay independently of each other.

We do not know n or p.

We do know that np = 1.4.

\[ X \sim \text{Bin}(n, p) \]

The number of decays in next hour is \( X \sim \text{Bin}(n, p) \).
Since $n$ is very large, we may approx.
the number of decays $X$ by $X' \sim \text{Poisson}(1.4)$.

$$\Rightarrow \quad P(X < 3) \approx P(X' < 3)$$

\[ P(X = 0) + P(X = 1) + P(X = 2) \]
\[ \downarrow \quad \downarrow \quad \downarrow \]
\[ P(X' = 0) \quad P(X' = 1) \quad P(X' = 2) \]

\[ P(X < 3) \approx P(X' < 3) = P(X' = 0) + P(X' = 1) + P(X' = 2) \]
\[ = e^{-1.4} + e^{-1.4} \cdot 1.4 + e^{-1.4} \cdot \frac{1.4^2}{2!} \approx 83\% \]

Midterm material ends here!