

Statistical Theory and Modeling (ST2601)

Discrete random variables

Mattias Villani

**Department of Statistics
Stockholm University**



mattiasvillani.com



@matvil



@matvil



mattiasvillani

Overview

- Random variables recap
- Bernoulli, Geometric and Binomial distributions
- Negative binomial distribution
- Chebychev's inequality

Probabilities of events

- **Probabilities** for events A and B in a **sample space** S .

$$0 \leq \Pr(A) \leq 1$$

- **Complement rule**

$$\Pr(\underbrace{A^c}_{\text{not } A}) = 1 - \Pr(A)$$

- **Addition rule**

$$\Pr(\underbrace{A \cup B}_{\text{union}}) = \Pr(A) + \Pr(B) - \Pr(\underbrace{A \cap B}_{\text{intersection}})$$













- **Multiplication rule**













$$\Pr(A \cap B) = \underbrace{\Pr(A|B)}_{\text{conditional prob}} \cdot \Pr(B)$$













- Multiplication rule when A and B are **independent**

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$













Throwing two dice

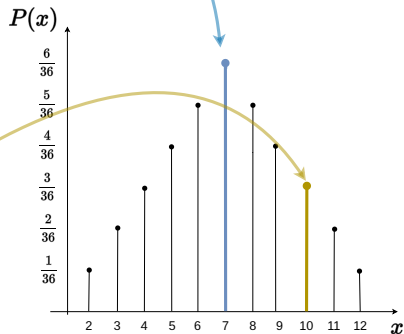
						
	2	3	4	5	6	7
	3	4	5	6	7	8
	4	5	6	7	8	9
	5	6	7	8	9	10
	6	7	8	9	10	11
	7	8	9	10	11	12

						
	2	3	4	5	6	7
	3	4	5	6	7	8
	4	5	6	7	8	9
	5	6	7	8	9	10
	6	7	8	9	10	11
	7	8	9	10	11	12

						
	2	3	4	5	6	7
	3	4	5	6	7	8
	4	5	6	7	8	9
	5	6	7	8	9	10
	6	7	8	9	10	11
	7	8	9	10	11	12

Random variables and probability distributions

						
	2	3	4	5	6	7
	3	4	5	6	7	8
	4	5	6	7	8	9
	5	6	7	8	9	10
	6	7	8	9	10	11
	7	8	9	10	11	12



Mean and variance

- **Discrete variable** with support $x \in \{x_1, x_2, \dots, x_K\}$ and

$$p_k = \Pr(X = x_k)$$

- **Expected value (mean)** is the **center** of the distribution

$$\mathbb{E}(X) = \sum_{k=1}^K x_k \cdot p_k$$

- Alternative: **probability function** $p(x)$

$$\mathbb{E}(X) = \sum_x x \cdot p(x)$$

where the sum implicitly is over all $x \in \{x_1, x_2, \dots, x_K\}$.

- **Variance** measures the **spread** of the distribution

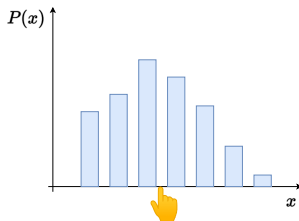
$$\sigma^2 = \mathbb{V}(X) = \mathbb{E}((X - \mu)^2) = \mathbb{E}(X^2) - \mu^2$$

- **Standard deviation** (same units as X)

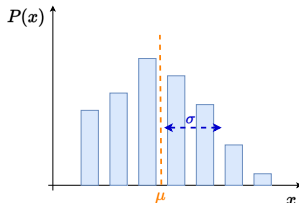
$$\sigma = \mathbb{S}(X) = \sqrt{\mathbb{V}(X)}$$

Mean and variance

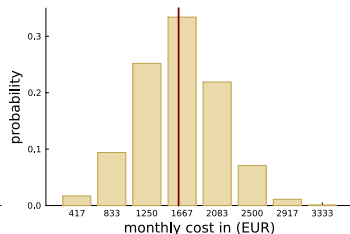
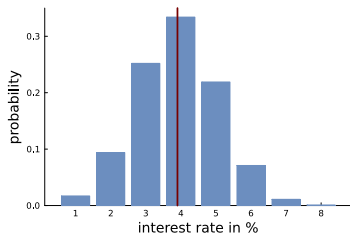
- The **mean** is where the probability distribution **balances**



- The **standard deviation** measures the **spread around μ** .



Example: Taking a 500,000 Euro bank loan



■ Mean interest rate

$$1 \cdot 0.017 + 2 \cdot 0.094 + \dots + 8 \cdot 0.001 \approx 3.9\%$$

■ Mean monthly cost for a 500000 Euro loan:

$$\mathbb{E}(\text{cost}) = 417 \cdot 0.017 + 833 \cdot 0.094 + \dots + 3333 \cdot 0.001 \approx 1626 \text{ EUR}$$

■ Variance monthly cost (in Euro²)

$$\mathbb{V}(\text{cost}) = (417 - 1626)^2 \cdot 0.017 + \dots + (3333 - 1626)^2 \cdot 0.001 \approx 241368$$

■ Standard deviation monthly cost

$$\mathbb{S}(\text{cost}) = \sqrt{241368} \approx 491 \text{ EUR}$$

Law of the unconscious statistician

- Let $g(Y)$ be a **function of the random variable** Y .
- The function need **not** be one-to-one.
- Theorem 3.2 in the WMS book

$$\mathbb{E}(g(Y)) = \sum_{\text{ally}} g(y) \cdot p(y)$$

- This result allows us to compute the mean of the new random variable $g(Y)$ *without computing its probability distribution*.
- Unconscious, since we do it almost without thinking.

Mean and variance of a linear transformation

Mean and variance of a linear transformation

Shift with constant c

$$\mathbb{E}(X + c) = \mathbb{E}(X) + c$$

$$\mathbb{V}(X + c) = \mathbb{V}(X)$$

Scaling with constant a

$$\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}(X)$$

$$\mathbb{V}(a \cdot X) = a^2 \mathbb{V}(X)$$

Linear transformation

$$\mathbb{E}(c + a \cdot X) = c + a \cdot \mathbb{E}(X)$$

$$\mathbb{V}(c + a \cdot X) = a^2 \mathbb{V}(X)$$

Mean and variance of a sum

Mean and variance of a sum of independent variables

If X and Y are independent random variables, then

Sum of two random variables

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$$

Linear transformation

$$\mathbb{E}(a \cdot X + b \cdot Y) = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y)$$

$$\mathbb{V}(a \cdot X + b \cdot Y) = a^2 \mathbb{V}(X) + b^2 \mathbb{V}(Y)$$

If X_1, \dots, X_n are independent random variables, then

Sum of n random variables

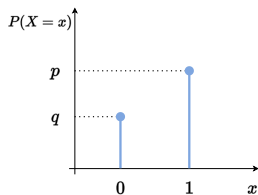
$$\mathbb{E}(X_1 + \dots + X_n) = \mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)$$

$$\mathbb{V}(X_1 + \dots + X_n) = \mathbb{V}(X_1) + \dots + \mathbb{V}(X_n)$$

Bernoulli distribution

- Success/Failure. $X \in \{0, 1\}$
- $X \sim \text{Bernoulli}(p)$, where p is success probability.
- **Probability function**

$$p(x) = \begin{cases} p & \text{for } x = 1 \\ q = 1 - p & \text{for } x = 0 \end{cases}$$



- **Mean and Variance**

$$\mathbb{E}(X) = p$$

$$\mathbb{V}(X) = pq$$

Binomial distribution

- $X_1, X_2, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Bernoulli}(p)$
- Then $Y = X_1 + X_2 + \dots + X_n$ follows a binomial distribution

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

- **Mean** and **Variance**

$$\mathbb{E}(X) = np$$

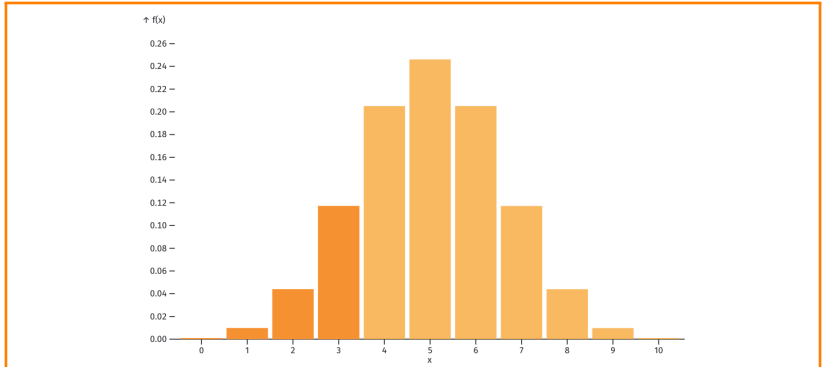
$$\mathbb{V}(X) = npq$$

- Proof: use that binomial = sum of independent Bernoullis.
- **Probability function**

$$p(x) = \binom{n}{x} p^x q^{n-x}$$

- Binomial does not care about the order, so $(0, 1, 1) = (1, 0, 1)$ etc. The **binomial coefficient** $\binom{n}{x}$ counts the number of ways we can order x successes in n trials.

Binomial distribution - widget



Geometric distribution - total number of trials

- Counts the number of Bernoulli trials **until first success**.
- $X \sim \text{Geom}(p)$ where $X \in \{1, 2, \dots\}$ and

$$p(x) = \Pr(\text{first success on trial } x) = \overbrace{q \cdot q \cdots q}^{x-1 \text{ failures}} \cdot \underbrace{p}_{\text{success}} = q^{x-1} \cdot p$$

multiply because indep

- This definition is used in the course book.
- **Mean** and **Variance**

$$\mathbb{E}(X) = \frac{1}{p}$$
$$\mathbb{V}(X) = \frac{1-p}{p^2}$$

- Proof involves the geometric series $\sum_{k=1}^{\infty} q^k = \frac{q}{1-q}$.

Geometric distribution - number of failures

- This an **alternative definition** of the Geometric distribution.
- Counts the **number of failed** trials **before first success**.
- Wikipedia has both definitions side-by-side.
- R, my widget and the formula sheet uses the one on this slide.
- $X \sim \text{Geom}(p)$ where $X \in \{0, 1, 2, \dots\}$ and

$$p(x) = \Pr(x \text{ failures before first success}) = \overbrace{q \cdot q \cdots q \cdot}^{x \text{ failures}} \underbrace{p}_{\text{success}} = q^x \cdot p$$

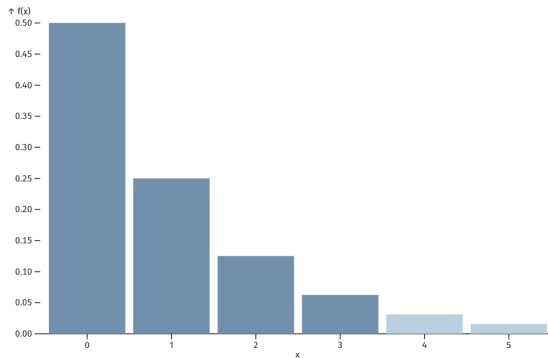
multiply because indep

- **Mean** and **Variance**

$$\mathbb{E}(X) = \frac{1 - p}{p}$$

$$\mathbb{V}(X) = \frac{1 - p}{p^2}$$

Geometric distribution - widget



Poisson distribution

- $X \sim \text{Pois}(\lambda)$ where $X \in \{0, 1, 2, \dots\}$

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

- Approximates $\text{Bin}(n, p)$ distribution for large n and small p .
- **Mean** and **Variance**

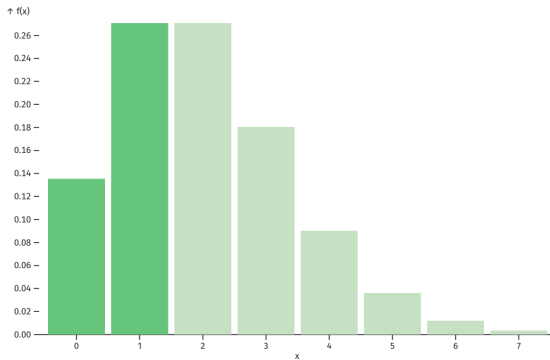
$$\mathbb{E}(X) = \lambda$$

$$\mathbb{V}(X) = \lambda$$

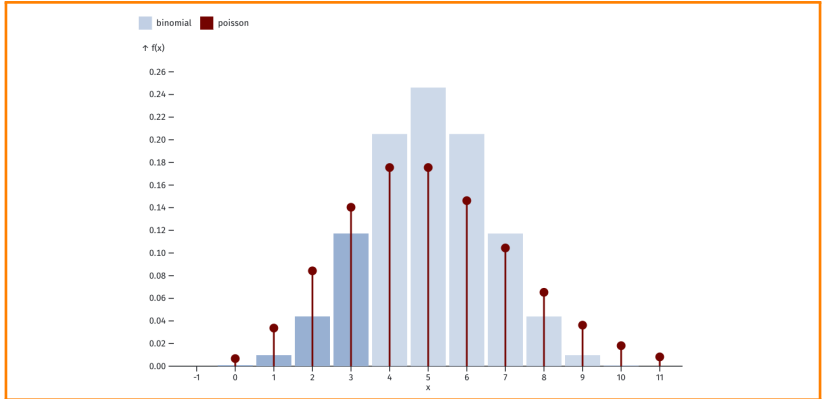
- **Mean = Variance.** Can be restrictive for real data.
- Proofs involve (see Taylor approximation in prequel if curious)

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Poisson distribution - widget



Poisson approximates Binomial - widget



Negative binomial distribution

- $X =$ **total number of trials** until r successes
- Total = failures + successes
- $X \sim \text{NegBin}(r, p)$ where $X \in \{r, r+1, r+2, \dots\}$

$$p(x) = \binom{x-1}{r-1} p^r q^{x-r}$$

- **Mean** and **Variance**

$$\mathbb{E}(X) = \frac{r}{p} \qquad \mathbb{V}(X) = \frac{r(1-p)}{p^2}$$

- Alternatively, count $X =$ **number of failures** before r successes. Then $X \in \{0, 1, 2, \dots\}$ and

$$\mathbb{E}(X) = \frac{r(1-p)}{p} \qquad \mathbb{V}(X) = \frac{r(1-p)}{p^2}$$

This is used in R, see the help `?dnbinom`

Negative binomial - mean parameterization

- Parameters p and r come naturally from Bernoulli trials.
- When modeling data, more interpretable to use:
 - ▶ $X =$ **number of failures**, and
 - ▶ parameterization $\text{NegBin}(r, \mu)$ with the **mean μ as an explicit parameter**.

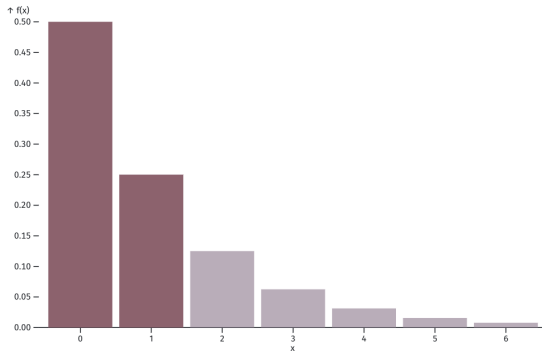
- Set $p = \frac{r}{r+\mu}$. Then, $\mathbb{E}(X) = \mu$, so μ is really the mean.
- The variance is

$$\mathbb{V}(X) = \frac{r(1-p)}{p^2} = \frac{\mu}{p} = \frac{\mu}{\left(\frac{r}{r+\mu}\right)} = \mu \left(1 + \frac{\mu}{r}\right)$$

so smaller r gives larger variance.

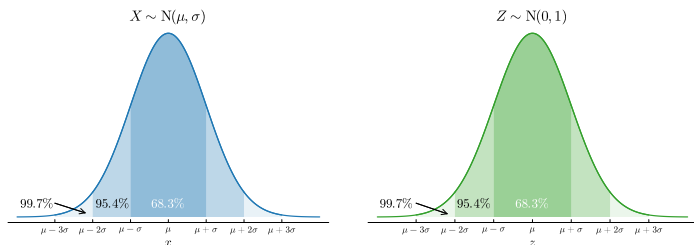
- The parameter r models **overdispersion** $\mathbb{V}(X) > \mathbb{E}(X)$.
We can let r be any positive real number, not just an integer.
- As $r \rightarrow \infty$, $\text{NegBin}(r, \mu)$ becomes $\text{Pois}(\mu)$.

Negative binomial distribution - widget



Chebyshev's inequality

Normal distribution 68-95-99.7% rule



Chebyshev: for **any** distribution with mean μ and variance σ^2

$$\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Chebyshev's bound is usually not tight:

- ▶ Normal: $\Pr(|X - \mu| \geq 2\sigma) \approx 0.0455$
- ▶ Chebshev: $\Pr(|X - \mu| \geq 2\sigma) \leq \frac{1}{2^2} = 0.25$

Useful for proofs, however.

Chebyshev's inequality - widget

True probability: $\Pr(|Y - \mu| \geq 2.18\sigma) = 0.03949$

Chebyshev's bound: $\Pr(|Y - \mu| \geq 2.18\sigma) \leq \frac{1}{2.18^2} = 0.21042$

