Statistical Theory and Modeling (ST2601)
Integration

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**

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

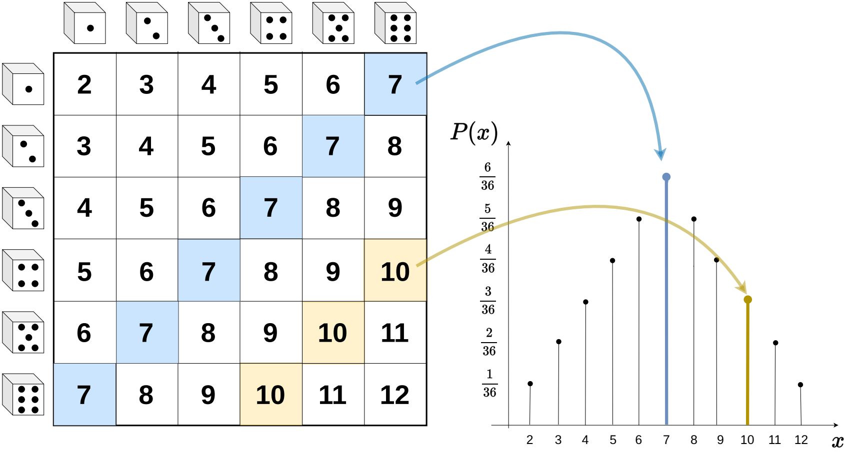
Throwing two dice

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	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



Mean and variance

- **Discrete variable** with support $x \in \{x_1, x_2, \ldots, x_K\}$ 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 implicity is over all $x \in \{x_1, x_2, \ldots, x_K\}$.

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

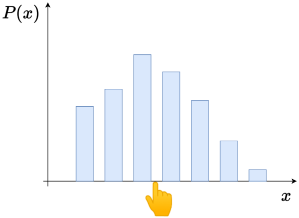
$$\sigma^2 = \mathbb{V}(X) = \mathbb{E}\left((X - \mu)^2\right) = \mathbb{E}(X^2) - \mu^2$$

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

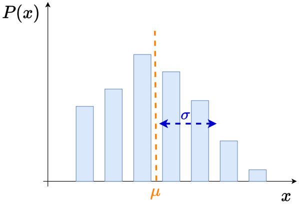
$$\sigma = \mathbb{S}(X) = \sqrt{\mathbb{V}(X)}$$

Mean and variance

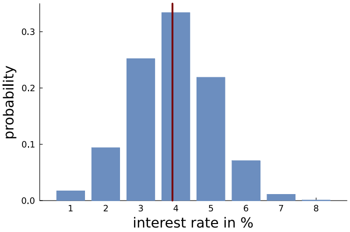
- The **mean** is where the probability distribution **balances**

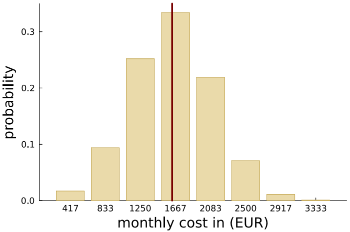


- The **standard deviation** measures the **spread around μ**.



Example: Taking a 500,000 Euro bank loan





- Mean interest rate

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

- **Mean monthly cost** for a 500000 Euro loan:

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

- **Variance monthly cost** (in Euro^2)

  $\mathbb{V}(\text{cost}) = (417 - 3252)^2 \cdot 0.017 + \ldots + (3333 - 1626)^2 \cdot 0.001 \approx 241368$

- **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

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

- 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 linear transformation

Shift with constant c

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

Scaling with constant a

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

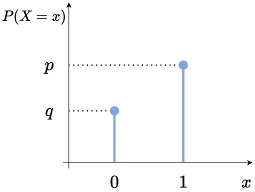
Linear transformation

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

Bernoulli distribution

- Success/Failure. $X \in \{0, 1\}$
- $X \sim \mathrm{Bernoulli}(p)$, 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

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

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

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

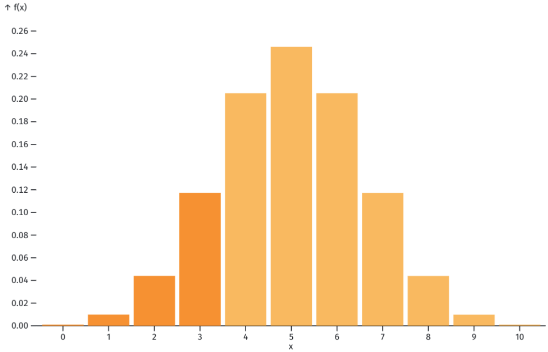
$$\mathbb{E}(X) = np$$
$$\mathbb{V}(X) = npq$$

- 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

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

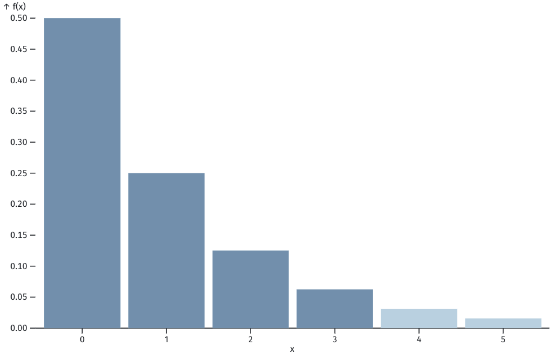
$$p(x) = \Pr(\text{first success on trial } x) = \overbrace{\underbrace{q \cdot q \cdots q}_{x-1 \text{ failures}} \cdot \underbrace{p}_{\text{success}}}^{\text{multiply because indep}} = q^{x-1} \cdot p$$

- Careful: sometimes $X =$ number of failures until first success. For example in my widget. Then $X \in \{0, 1, \ldots\}$.
- **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 - widget



Poisson distribution

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

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

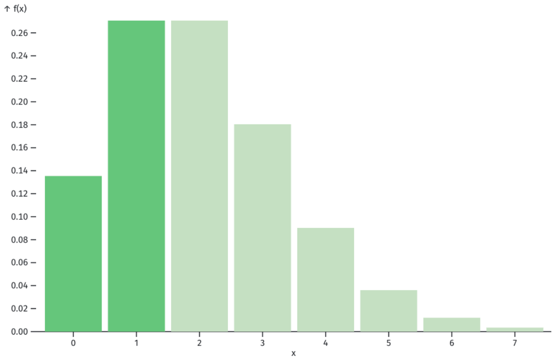
- Approximates $\mathrm{Bin}(n, p)$ distribution for large n and small p.
- **Mean** and **Variance**

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

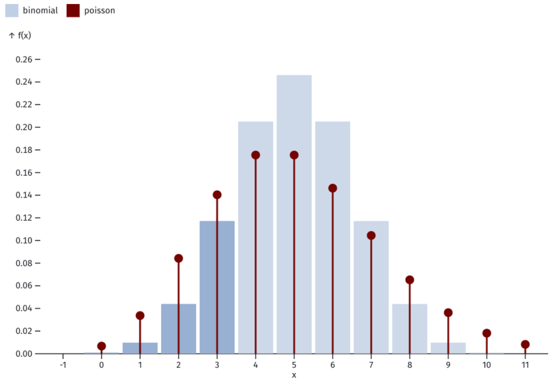
- **Mean = Variance**. Can be restrictive for real data.
- Proofs involve (see Taylor approximation in prequel if curious)

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

Poisson distribution - widget



Poisson approximates Binomial - widget



Negative binomial distribution

- Number of trials until r successes.
- $X \sim \mathrm{NegBin}(r, p)$ where $X \in \{r, r+1, r+2, \ldots\}$

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

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

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

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

- Alternative def 1: counts the number of failures before r successes. Then $X \in \{0, 1, 2, ...\}$.
- Alternative def 2: use mean as parameter. $X \sim \mathrm{NegBin}(\lambda, \phi)$

$$\mathbb{E}(X)\mathbb{V}(X) = \lambda \qquad = \lambda\left(1 + \frac{\lambda}{\phi}\right)$$

- As $\phi \to \infty$. Becomes $\mathrm{Pois}(\lambda)$. ϕ models **overdispersion**.

Negative binomial distribution - widget