Probability and Statistics

3 – Discrete Random Variables

Stefan Heiss

Technische Hochschule Ostwestfalen-Lippe
Dep. of Electrical Engineering and Computer Science

November 10, 2023

Skewness and Kurtosis

Definition (3.39)

Let X be a random variable with $m = E(X)$ and $\sigma^2 = Var(X) > 0$. The third and fourth moment of the normalized random variable defined by X are the *skewness* and *kurtois* of X:

$$\mathrm{skew}(X) \;:=\; E\left(\left(\frac{X-m}{\sigma}\right)^3\right)$$

$$\mathrm{kurt}(X) \;:=\; E\left(\left(\frac{X-m}{\sigma}\right)^4\right)$$

The skewness measures the lack of symmetry, while the kurtosis measures the fatness in the tails of the pdf of X.

Skewness and Kurtosis

Lemma (3.40)

(i) $$\operatorname{skew}(X) = \frac{E\left(X^3\right) - 3m\sigma^2 - m^3}{\sigma^3}$$

(ii) $$\operatorname{kurt}(X) = \frac{E\left(X^4\right) - 4mE\left(X^3\right) + 6m^2\sigma^2 + 3m^4}{\sigma^4}$$

Markov's inequality

Theorem (3.41)

If X is a nonnegative random variable and $a > 0$, then:

$$\Pr(X \geq a) \quad \leq \quad \frac{E(X)}{a}$$

Pf.:

$$E(X) = \sum_{x \in \mathbb{R}^+} x \cdot p_X(x) \geq \sum_{x \geq a} x \cdot p_X(x) \geq \sum_{x \geq a} a \cdot p_X(x) = a \cdot \sum_{x \geq a} p_X(x)$$

$$= a \cdot \Pr(X \geq a)$$

$$\frac{E(X)}{a} \geq \Pr(X \geq a) \quad \square$$

Chebyshev's inequality

Theorem (3.42)

If X is a random variable and $a > 0$, then: $\Pr\left(|X| \geq a\right) \leq \dfrac{E(X^2)}{a^2}$ (i)

Furthermore, if $m = E(X)$ is finite, then: $\Pr\left(|X - m| \geq a\right) \leq \dfrac{Var(X)}{a^2}$ (ii)

(i) $\Pr(|X| \geq a) = \Pr(X^2 \geq a^2) \overset{(3.41)}{\leq} E(X^2)/a^2$ ✓

(ii) $\Pr(|X-m| \geq a) \overset{(i)}{\leq} \dfrac{E\left((X-m)^2\right)}{a^2} = \dfrac{Var(X)}{a^2}$ ✓

Weak law of large numbers

Theorem (3.43)

Let $X_1, X_2, \ldots$ be a sequence of uncorrelated random variables with a common mean $m = E(X_i)$ and a common variance $\sigma^2 = Var(X_i)$ for all $i \in \mathbb{N}$. For every $n \in \mathbb{N}$ the sample mean *of the first n X_i's is defined by:*

$$M_n := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$$E(M_n) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} n \cdot m = m$$

Then for any $\varepsilon > 0$:

$$\Pr\left(|M_n - m| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\,\varepsilon^2} \longrightarrow 0 \qquad \text{for } n \to \infty$$

Proof of (3.43)

$$
\begin{aligned}
\Pr\left(|M_n - m| \geq \varepsilon\right) &= \Pr\left(|M_n - E(M_n)| \geq \varepsilon\right) && \checkmark \\
&\leq \frac{1}{\varepsilon^2}\, Var(M_n) && \text{(3.42)(ii)} \\
&= \frac{1}{n^2\,\varepsilon^2}\, Var\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2\,\varepsilon^2} \sum_{i=1}^{n} Var(X_i) && X_i \text{ are uncorollated} \\
&= \frac{\sigma^2}{n\,\varepsilon^2}
\end{aligned}
$$

Weak law of large numbers

$X_i \sim \text{Bernoulli}(1/2)$

$M_n = \frac{1}{n}(X_1 + \dots + X_n)$

Example (3.44)

How often must a fair coin be flipped such that, with a probability $p \geq 0.98$, "heads" show up in 49%–51% of all cases?

$$\Pr(0.49 \leq M_n \leq 0.51) \geq 0.98$$

$$\Pr(|M_n - 0.5| \geq 0.01) \leq 0.02$$

$$\wedge|$$

holds if: $\frac{\sigma^2}{n\varepsilon^2} = \frac{1/4}{n \cdot (0.01)^2} \leq 0.02$

$$\Updownarrow$$

$$n \geq \frac{1}{4} \cdot 10{,}000 \cdot \frac{100}{2} = 125.000$$

Selected Discrete Probability Distributions

- Uniform Distributions
- Bernoulli Distributions
- Binomial Distributions
- Geometric Distributions
- Negative Binomial Distributions
- Poisson Distributions
- Hypergeometric Distributions

Poisson approximation of binomial probabilities

For small values of p ($p \lesssim 0.01$), the binomial distribution can be approximated by the Poisson distribution with mean $\lambda = np$.

To be more accurate:

Theorem (3.59)

If $(p_n)_{n\in\mathbb{N}}$ is a sequence of numbers with $p_n \in (0,1)$ and

$$\lim_{n\to\infty} n \cdot p_n \;=\; \lambda$$

for some $\lambda \in \mathbb{R}$, then the sequence of binomial distributions $\mathrm{binomial}(n, p_n)$ converges to the Poisson distribution $\mathrm{Poisson}(\lambda)$.

Proof of (3.59)

$p_n = \frac{\lambda}{n}$

$$
\begin{aligned}
p_{\mathrm{binomial}(n,p_n)}(i) &= \binom{n}{i} \cdot p_n^i (1-p_n)^{n-i} \\
&= \frac{n(n-1)\cdots(n-(i-1))}{i!} \cdot p_n^i (1-p_n)^{n-i} \\
&= \frac{(np_n)^i}{i!} \cdot \frac{n(n-1)\cdots(n-(i-1))}{n^i} \cdot (1-p_n)^{n-i} \\
&= \frac{(\overset{\lambda}{np_n})^i}{i!} \cdot \underset{\searrow 1}{\left(1-\frac{1}{n}\right)} \cdots \underset{\searrow 1}{\left(1-\frac{i-1}{n}\right)} \cdot \underbrace{(1-\overset{0}{p_n})}_{\searrow 1}{}^{-\overset{\text{fixed}}{i}} \underbrace{\left(1-\frac{\overset{\lambda}{np_n}}{n}\right)^n}_{\hookrightarrow e^{-\lambda}} \\
&\to \frac{\lambda^i}{i!} \cdot e^{-\lambda} \qquad \text{for } n \to \infty
\end{aligned}
$$

$P_{\mathrm{Poisson}(\lambda)}(i)$

Poisson approximation of binomial probabilities

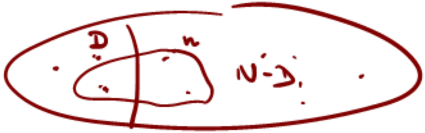
Theorem (3.60)

Let $n \in \mathbb{N}$ *and* $p \in (0,1)$. *Then:*

$$\sum_{i=0}^{\infty} \left| p_{\mathrm{binomial}(n,p)}(i) - p_{\mathrm{Poisson}(n \cdot p)}(i) \right| \leq 2np^2$$

Hypergeometric Random Variables

Assume, a total number of N entities are given containing D defective elements. Furthermore assume that $n \leq N$ elements are drawn randomly without replacement. Let $X_i = 1$ if the i'th element drawn is defective and $X_i = 0$ otherwise. Then $X = X_1 + \cdots + X_n$ is a hypergeometric random variable:



$n \leq N$

Hypergeometric Random Variables

Assume, a total number of N entities are given containing D defective elements. Furthermore assume that $n \leq N$ elements are drawn randomly without replacement. Let $X_i = 1$ if the i'th element drawn is defective and $X_i = 0$ otherwise. Then $X = X_1 + \cdots + X_n$ is a hypergeometric random variable:

Definition (3.62)

A *hypergeometric random variable* is a random variable X having a distribution given by

$$p_X(i) := \frac{\binom{D}{i}\binom{N-D}{n-i}}{\binom{N}{n}} \qquad \text{for } i = 0, 1, \ldots, n,$$

← # of all possible samples of size n

where $N, D, n \in \mathbb{N}$ are fixed parameters with $D \leq N$ and $n \leq N$. This may be denoted by $X \sim \mathrm{hypergeometric}(N, D, n)$.

Hypergeometric Random Variables

Lemma (3.63)

If $X \sim \mathrm{hypergeometric}(N, D, n)$ *and* $p := \frac{D}{N}$, *then:*

(i) $$E(X) = \frac{n\,D}{N} = np$$

(ii) $$Var(X) = \frac{nD(N-D)}{N^2}\left(1 - \frac{n-1}{N-1}\right) = np(1-p)\cdot\left(1 - \frac{n-1}{N-1}\right)$$

Proof of (3.63)

X_i's are not independent

(i) $$E(X) = E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = n\,\frac{D}{N}$$

Proof of (3.63)

(i)

$$E(X) \;=\; E(X_1 + \cdots + X_n) \;=\; E(X_1) + \cdots + E(X_n) \;=\; n\,\frac{D}{N}$$

(ii)

$$\begin{aligned} Var(X) &= \sum_{i=1}^{n} Var(X_i) \;+\; 2 \cdot \sum_{1\le i<j\le n} Cov(X_i, X_j) \\ &= n \cdot \frac{D}{N} \cdot \frac{N-D}{N} \;+\; 2 \cdot \sum_{1\le i<j\le n} Cov(X_i, X_j) \end{aligned}$$

Proof of (3.63)

(i)

$$E(X) = E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = n\,\frac{D}{N}$$

(ii)

$$\begin{aligned} Var(X) &= \sum_{i=1}^{n} Var(X_i) + 2 \cdot \sum_{1 \le i < j \le n} Cov(X_i, X_j) \\ &= n \cdot \frac{D}{N} \cdot \frac{N-D}{N} + 2 \cdot \sum_{1 \le i < j \le n} Cov(X_i, X_j) \end{aligned}$$

For $i \neq j$, $X_i \cdot X_j$ is a Bernoulli variable with:

$$\Pr(X_i \cdot X_j = 1) = \frac{D(D-1)}{N(N-1)}$$

Proof of (3.63)

Therefore

$$\begin{aligned} Cov(X_i, X_j) &= E(X_iX_j) - E(X_i)E(X_j) \\ &= \frac{D(D-1)}{N(N-1)} - \left(\frac{D}{N}\right)^2 = \frac{-D(N-D)}{N^2(N-1)} \end{aligned}$$

Proof of (3.63)

Therefore

$$
\begin{aligned}
Cov(X_i, X_j) &= E(X_i X_j) - E(X_i)E(X_j) \\
&= \frac{D(D-1)}{N(N-1)} - \left(\frac{D}{N}\right)^2 = \frac{-D(N-D)}{N^2(N-1)}
\end{aligned}
$$

and it follows that:

$$
Var(X) = \frac{n\,D\,(N-D)}{N^2} - 2\binom{n}{2}\frac{D(N-D)}{N^2(N-1)} = \frac{n\,D\,(N-D)}{N^2}\left(1 - \frac{n-1}{N-1}\right)
$$