Introduction to probability

Stat 134       Fall 2005
Berkeley

Lectures prepared by:
Elchanan Mossel
elena Shvets

Follows Jim Pitman’s book:
Probability
Section 3.3
$X = 2 \times \text{Bin}(300, 1/2) - 300$

$E[X] = 0$
Histo 2

\[ Y = 2 \times \text{Bin}(30, 1/2) - 30 \]

\[ E[Y] = 0 \]
$Z = 4 \times \text{Bin}(10, 1/4) - 10$

$E[Z] = 0$
\( W = 0 \)
\( E[W] = 0 \)
A natural question:

• Is there a good parameter that allow to distinguish between these distributions?
• Is there a way to measure the spread?
Variance and Standard Deviation

• The variance of $X$, denoted by $\text{Var}(X)$ is the mean squared deviation of $X$ from its expected value $\mu = E(X)$:

$$\text{Var}(X) = E[(X-\mu)^2].$$

The standard deviation of $X$, denoted by $\text{SD}(X)$ is the square root of the variance of $X$:

$$\text{SD}(X) = \sqrt{\text{Var}(X)}.$$
Computational Formula for Variance

Claim:

\[
\text{Var}(X) = E(X^2) - E(X)^2.
\]

Proof:

\[
E[(X - \mu)^2] = E[X^2 - 2\mu X + \mu^2]
\]

\[
E[X^2] - 2\mu E[X] + \mu^2
\]

\[
E[X^2] - 2\mu^2 + \mu^2
\]

\[
E[X^2] - E[X]^2
\]
Properties of Variance and SD

1. **Claim**: \( \text{Var}(X) \geq 0. \)

   **Pf**: \( \text{Var}(X) = \sum (x-\mu)^2 \cdot P(X=x) \geq 0 \)

2. **Claim**: \( \text{Var}(X) = 0 \iff P[X=\mu] = 1. \)
Variance and SD

For a general distribution Chebyshev inequality states that for every random variable $X$, $X$ is expected to be close to $E(X)$ give or take a few $SD(X)$.

Chebyshev Inequality:

For every random variable $X$ and all $k > 0$:

$$P(|X - E(X)| \geq k \cdot SD(X)) \leq \frac{1}{k^2}.$$
Chebyshev’s Inequality

\[ P(\lvert X - E(X)\rvert \geq k \ SD(X)) \leq \frac{1}{k^2} \]

**proof:**

- Let \( \mu = E(X) \) and \( \sigma = SD(X) \).
- Observe that \( \lvert X-\mu\rvert \geq k \sigma \iff \lvert X-\mu\rvert^2 \geq k^2 \sigma^2 \).
- The RV \( |X-\mu|^2 \) is non-negative, so we can use Markov’s inequality:
- \( P(|X-\mu|^2 \geq k^2 \sigma^2) \leq E [ |X-\mu|^2 ] / k^2 \sigma^2 \)
- \( P(|X-\mu|^2 \geq k^2 \sigma^2) \leq \sigma^2 / k^2 \sigma^2 = 1/k^2. \)
Variance of Indicators

Suppose $I_A$ is an indicator of an event $A$ with probability $p$. Observe that $I_A^2 = I_A$.

\[
E(I_A^2) = E(I_A) = P(A) = p, \text{ so: } Var(I_A) = E(I_A^2) - E(I_A)^2 = p - p^2 = p(1-p).
\]
Variance of a Sum of Independent Random Variables

Claim: if $X_1, X_2, \ldots, X_n$ are independent then:

$$\text{Var}(X_1 + X_2 + \ldots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \ldots + \text{Var}(X_n).$$

Pf: Suffices to prove for 2 random variables.

$$E[(X+Y - E(X+Y))^2] = E[(X-E(X) + Y-E(Y))^2] =$$

$$E[(X-E(X))^2] + 2E[(X-E(X))(Y-E(Y))] + E(Y-E(Y))^2] =$$

$$\text{Var}(X) + \text{Var}(Y) + 2E[(X-E(X)) E(Y-E(Y))] \ (\text{mult.rule}) =$$

$$\text{Var}(X) + \text{Var}(Y) + 0$$
Variance and Mean under scaling and shifts

- **Claim:** SD(aX + b) = |a| SD(X)
- **Proof:**
  \[
  \text{Var}[aX+b] = \text{E}[(aX+b - a\mu -b)^2] = \\
  = \text{E}[a^2(X-\mu)^2] = a^2 \sigma^2
  \]
- **Corollary:** If a random variable X has
  - E(X) = \mu and SD(X) = \sigma > 0, then
  - \(X^*=(X-\mu)/\sigma\) has
  - E(X^*) = 0 and SD(X^*)=1.
Square Root Law

Let $X_1, X_2, \ldots, X_n$ be independent random variables with the same distribution as $X$, and let $S_n$ be their sum:

$$S_n = \sum_{i=1}^{n} X_i,$$
and their average, then:

$$\bar{X} = \frac{S_n}{n}$$

$$E(S_n) = nE(X)$$

$$Var(S_n) = nVar(X)$$

$$SD(S_n) = \sqrt{n}SD(X).$$

$$E(\bar{X}_n) = E(X)$$

$$SD(\bar{X}_n) = \frac{SD(X)}{\sqrt{n}}$$
Weak Law of large numbers

**Thm:** Let $X_1, X_2, \ldots$ be a sequence of independent random variables with the same distribution. Let $\mu$ denote the common expected value

\[ \mu = E(X_i). \]

And let

\[ \bar{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}. \]

Then for every $\varepsilon > 0$:

\[ P(|\bar{X}_n - \mu| < \varepsilon) \rightarrow 1 \text{ as } n \rightarrow \infty. \]
Weak Law of large numbers

Proof: Let $\mu = \text{E}(X_i)$ and $\sigma = \text{SD}(X_i)$. Then from the square root law we have:

$$E(\bar{X}_n) = \mu \text{ and } \text{SD}(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}.$$ 

Now Chebyshev inequality gives us:

$$P(|\bar{X}_n - \mu| \geq \varepsilon) = P(|\bar{X}_n - \mu| \geq \frac{\varepsilon\sqrt{n}}{\sigma} \cdot \frac{\sigma}{\sqrt{n}}) \leq \left(\frac{\sigma}{\varepsilon\sqrt{n}}\right)^2$$

For a fixed $\varepsilon$ right hand side tends to 0 as $n$ tends to $\infty$. 
The Normal Approximation

• Let $S_n = X_1 + \ldots + X_n$ be the sum of independent random variables with the same distribution.

• Then for large $n$, the distribution of $S_n$ is approximately normal with mean $E(S_n) = n \mu$ and $SD(S_n) = \sigma n^{1/2}$.

• where $\mu = E(X_i)$ and $\sigma = SD(X_i)$.

In other words:

$$P(a \leq \frac{S_n - n\mu}{\sigma \sqrt{n}} \leq b) \sim \Phi(b) - \Phi(a)$$
Sums of repeated independent random variables

Suppose $X_i$ represents the number obtained on the $i$'th roll of a die. Then $X_i$ has a uniform distribution on the set \{1,2,3,4,5,6\}. 
Distribution of $X_1$
Sum of two dice

We can obtain the distribution of $S_2 = X_1 + X_2$ by the convolution formula:

$$P(S_2 = k) = \sum_{i=1}^{k-1} P(X_1=i) P(X_2=k-i \mid X_1=i),$$

by independence

$$= \sum_{i=1}^{k-1} P(X_1=i) P(X_2=k-i).$$
Distribution of $S_2$
Sum of four dice

We can obtain the distribution of

\[ S_4 = X_1 + X_2 + X_3 + X_4 = S_2 + S'_2 \]
again by the convolution formula:

\[
P(S_4 = k) = \sum_{i=1}^{k-1} P(S_2 = i) \cdot P(S'_2 = k - i \mid S_2 = i),
\]
by independence of \( S_2 \) and \( S'_2 \)

\[
= \sum_{i=1}^{k-1} P(S_2 = i) \cdot P(S'_2 = k - i).
\]
Distribution of $S_4$
Distribution of $S_8$
Distribution of $S_{32}$
Distribution of $X_1$
Distribution of $S_2$
Distribution of $S_4$
Distribution of $S_8$
Distribution of $S_{16}$
Distribution of $S_{32}$
Distribution of $S_2$
Distribution of $S_4$
Distribution of $S_8$
Distribution of $S_{16}$
Distribution of $S_{32}$