# Lecture #2

- The expectation and variance have essential roles in statistics.
- They will often be the values we are interested in estimating, due to their intuitive meaning.
- We can also use them to control extreme behavior via concentration inequalities.

These typically bound deviations of a r.v. $X$ from its mean, or bound its tails, in terms of "simple" properties of $X$, i.e. its expectation and variance.

Theorem (Markov's Inequality): Let $X$ be a r.v. taking non-negative values. Suppose $\mathbb{E}(X)$ exists. Then $\forall t > 0$, $\mathbb{P}(X > t) \leq \dfrac{\mathbb{E}(X)}{t}$.

Proof: Assume WLOG that $X$ is continuous. Then we let $f_X(x)$ be the density of $X$ and note

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

Since $X$ takes non-negative values $\Rightarrow f_X(x) = 0$ for $x < 0$

$$= \int_{0}^{\infty} x f_X(x)\,dx$$

So, for a fixed $t > 0$, break up the integral at $t$ and analyze:

$$\mathbb{E}(X) = \int_{0}^{\infty} x f_X(x)\,dx$$

$$= \int_{0}^{t} x f_X(x)\,dx + \int_{t}^{\infty} x f_X(x)\,dx$$

$$\geq \int_t^\infty x f_X(x)\,dx$$

$$\geq t \int_t^\infty f_X(x)\,dx$$

$$= t \cdot \mathbb{P}(X \in [t,\infty))$$

$$\Rightarrow \mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}. \blacksquare$$

- Obviously, we threw away a lot. We tossed $\int_0^t x f_X(x)\,dx$, and also made the potentially very lax estimate $\int_t^\infty x f_X(x)\,dx$

$$\geq t \int_t^\infty f_X(x)\,dx.$$

- For these reasons, the inequality is often rather loose.
- Markov allows us to prove a result that depends on the variance of $X$.

Theorem (Chebyshev's Inequality): Let $X$ be a r.v. with finite expectation $\mathbb{E}(X)$ and finite variance $Var(X)$. Then $\forall t > 0$, $\mathbb{P}(|X - \mathbb{E}(X)| \geq t)$

$$\leq \frac{Var(X)}{t^2}.$$

<u>Proof</u>: We can show the desired result by applying Markov's inequality on $Y := |X - \mathbb{E}(X)|$. Then Markov gives, for any $t > 0$,

$$\begin{aligned} &\mathbb{P}(|X - \mathbb{E}(X)| > t) \\ &= \mathbb{P}(|X - \mathbb{E}(X)|^2 > t^2) \\ &\leq \frac{\mathbb{E}|X - \mathbb{E}(X)|^2}{t^2} \end{aligned}$$

But notice that $\mathbb{E}|X - \mathbb{E}(X)|^2$ is exactly $\mathrm{Var}(X)$. ∎

- Markov and Chebyshev are quite generic. In statistics, we will often be interested in computing <u>empirical</u> averages, so the following bound is often more useful.

  empirical $\to$ observed from data

<u>Theorem (Hoeffding)</u>: Let $\{X_i\}_{i=1}^n$ be independent r.v. that are

(i) Mean 0: $\mathbb{E}(X_i) = 0$, $i = 1, \dots, n$

(ii) Bounded: $a_i \leq X_i \leq b_i$, $i = 1, \dots, n$

Let $\varepsilon > 0$. Then for any $t > 0$,

$$\mathbb{P}\left(\sum_{i=1}^n X_i > \varepsilon\right) \leq \exp(-t\varepsilon) \cdot \prod_{i=1}^n \exp\left(t^2 [b_i - a_i]^2 / 8\right).$$