A Quick Guide to Markov, Chebyshev, Hoeffding, and Chernoff Inequalities

Setup. All random variables are defined on a common probability space. Write $\mathbb{E}$ for expectation, Var for variance, and $\mathbb{P}$ for probability. For sums, we often take $X = \sum_{i=1}^{n} X_i$ with the X_i independent.

1 Markov's Inequality

Statement. If $X \geq 0$ and $a > 0$, then

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Proof (one line). Since $X \geq a \cdot \mathbf{1}\{X \geq a\}$, taking expectations gives $\mathbb{E}[X] \geq a\,\mathbb{P}(X \geq a)$.

Use. Requires only nonnegativity and a finite mean. It's crude but universal; often used as a first step or applied to $g(X) \geq 0$ (e.g. $g(x) = e^{\lambda x}$).

2 Chebyshev's Inequality

Statement. For any $t > 0$ and any real X with finite mean μ and variance σ^2,

$$\mathbb{P}(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}.$$

Proof. Apply Markov to the nonnegative variable $(X-\mu)^2$: $\mathbb{P}\big((X-\mu)^2 \geq t^2\big) \leq \mathbb{E}[(X-\mu)^2]/t^2 = \sigma^2/t^2$.

Use. Tail bound that needs only variance; suboptimal constants but assumption-light.

3 Hoeffding's Inequality (Additive form)

Setting. $X_1, \ldots, X_n$ independent with $a_i \leq X_i \leq b_i$ a.s., and let $S = \sum_{i=1}^{n} X_i$, $\mu = \mathbb{E}[S]$.

Statement. For any $t > 0$,

$$\mathbb{P}\big(S - \mu \geq t\big) \;\leq\; \exp\Big(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\Big), \qquad \mathbb{P}\big(|S - \mu| \geq t\big) \;\leq\; 2\exp\Big(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\Big).$$

Proof sketch. For each i, by Hoeffding's lemma $\log \mathbb{E}\, e^{\lambda(X_i - \mathbb{E}X_i)} \le \lambda^2(b_i - a_i)^2/8$. Independence makes mgfs multiply; optimize over $\lambda > 0$.

Use. Concentration for *bounded* independent summands. Gives sub-Gaussian tails with variance proxy $\sum(b_i - a_i)^2/4$.

4 Chernoff Bounds (Multiplicative, for sums of Bernoulli)

Setting. $X = \sum_{i=1}^{n} X_i$ with $X_i \sim \text{Bernoulli}(p_i)$ independent; let $\mu = \mathbb{E}[X] = \sum p_i$.

Upper tail (multiplicative). For $0 < \delta < 1$,

$$\mathbb{P}\big(X \ge (1+\delta)\mu\big) \;\le\; \Big(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\Big)^{\mu} \;\le\; \exp\Big(-\frac{\mu\delta^2}{3}\Big).$$

For $\delta \ge 1$, $\mathbb{P}(X \ge (1+\delta)\mu) \le \exp\big(-\frac{\mu\delta}{3}\big)$.

Lower tail. For $0 < \delta < 1$,

$$\mathbb{P}\big(X \le (1-\delta)\mu\big) \;\le\; \Big(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\Big)^{\mu} \;\le\; \exp\Big(-\frac{\mu\delta^2}{2}\Big).$$

Proof sketch (Chernoff method). For $\lambda > 0$, $\mathbb{P}(X \ge t) = \mathbb{P}(e^{\lambda X} \ge e^{\lambda t}) \le e^{-\lambda t}\mathbb{E}e^{\lambda X}$. For Bernoulli, $\mathbb{E}e^{\lambda X} = \prod_i \big(1 - p_i + p_i e^{\lambda}\big)$. Optimize λ to get the Kullback–Leibler form; relax to the quadratic forms above.

Use. Sharpest classical tails for sums of independent indicators (or bounded r.v.'s via mgf bounds). Preferred when deviations scale relative to μ.

5 Connections and Quick Guidance

- **Markov** needs only $\mathbb{E}[X]$ and $X \ge 0$; weakest but universal.
- **Chebyshev** uses variance; symmetric two-sided bound, assumption-light.
- **Hoeffding** assumes bounded independent summands; gives sub-Gaussian tails in t with scale $\sqrt{n}$.
- **Chernoff** (for Bernoulli/Poisson-binomial) gives multiplicative bounds in μ; typically tighter than Hoeffding when p_i are small/moderate.

6 Mini Examples

Chebyshev sanity check

If X has mean μ and sd σ, then $\mathbb{P}(|X - \mu| \ge 3\sigma) \le 1/9 \approx 0.111$ (distribution-free).

Hoeffding for averages

Let $Y_i \in [0,1]$ independent, $\bar{Y} = \frac{1}{n}\sum Y_i$, and $t > 0$. Then

$$\mathbb{P}\big(|\bar{Y} - \mathbb{E}\bar{Y}| \geq t\big) \leq 2\exp(-2nt^2).$$

Chernoff for coin flips

$X \sim \mathrm{Bin}(n,p)$, $\mu = np$. For $0 < \delta < 1$,

$$\mathbb{P}\big(X \geq (1+\delta)\mu\big) \leq \exp\Big(-\frac{\mu\delta^2}{3}\Big), \quad \mathbb{P}\big(X \leq (1-\delta)\mu\big) \leq \exp\Big(-\frac{\mu\delta^2}{2}\Big).$$

7 One-Line Derivations to Remember

$$\begin{aligned}
\text{Markov: } \mathbb{P}(X \geq a) &\leq \frac{\mathbb{E}[X]}{a}. \\
\text{Chebyshev: } \mathbb{P}(|X - \mu| \geq t) &\leq \frac{\mathrm{Var}(X)}{t^2}. \\
\text{Chernoff trick: } \mathbb{P}(X \geq t) &\leq \inf_{\lambda>0}\, e^{-\lambda t}\, \mathbb{E}e^{\lambda X}. \\
\text{Hoeffding lemma: } \log \mathbb{E}e^{\lambda(X - \mathbb{E}X)} &\leq \frac{\lambda^2(b-a)^2}{8} \text{ if } X \in [a,b].
\end{aligned}$$