Probabilistic Tools

William Schultz

November 8, 2022

A *random variable* is a variable whose possible values are numerical outcomes of a random phenomenon. Each possible outcome is assigned some probability, with the condition that the sum of probabilities overall possible outcomes must sum to 1.

See [1] for a reference on some probabilistic tools.

Linearity of Expectation

For any two random variables X, Y, it holds that

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

Expectation and Variance

For a discrete random variable X taking values in some set $\Omega \subseteq \mathbb{R}$, its *expectation* is defined as

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} \omega \cdot Pr[X = \omega]$$

and its *variance* is defined as

$$\text{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

That is, expectation is essentially a weighted sum of the values that the random variable can take on, where each value is weighted by the probability of that event occurring, and variance is essentially a measure of the average deviation of the variable from its expectation/mean.

Markov's Inequality

Markov's inequality is an elementary large deviation bound valid for *all* non-negative random variables. Let X be a non-negative random variable with $\mathbb{E}[X] > 0$. Then, for all $\lambda > 0$

$$Pr(X \geq \lambda \mathbb{E}[X]) \leq \frac{1}{\lambda}$$

That is, this establishes a bound, for some given parameter λ, on how likely a random variable is to be far away from its mean. For example, if $\lambda = 10$, then this means that the probability that X is greater than 10 times its mean is $\leq \frac{1}{10}$. Note that this bound doesn't take into account anything about the actual distribution (only about its mean), so it may serve as only a rough estimate.

Chebyshev's Inequality

Let X be a random variable with $\text{Var}(X) > 0$, and where $\sigma = \sqrt{\text{Var}(X)}$ is its standard deviation. Then, for all $\lambda > 0$

$$Pr[|X - \mathbb{E}[X]| \geq \lambda \sigma] \leq \frac{1}{\lambda^2}$$

This bound tells us something about how far a random variable is from its expectation, in terms of the variance of the RV. That is, it puts a bound on the probability of how many (λ) standard deviations (σ) away from its mean X may be.

Chernoff Bound

Let $X_1, \ldots, X_n$ be independent random variables in $[0, 1]$. Let $X = \sum_{i=1}^{n} X_i$ and let $\mu = \mathbb{E}[X]$. Then

- **Upper Tail** $Pr[X \geq (1 + \delta)\mu] \leq \exp\left(-\frac{\delta^2}{2+\delta}\mu\right), \quad \forall \delta > 0$

- **Lower Tail**: $Pr[X \geq (1-\delta)\mu] \leq \exp\left(-\frac{\delta^2}{2}\mu\right), \quad \forall \delta \in [0,1]$

The two tail bounds above can be combined into the following (slightly weaker) two-sided tail bound:

- **Two-sided bound**: $Pr[|X - \mu| \geq \delta\mu] \leq 2\exp\left(-\frac{\delta^2\mu}{3}\right), \quad \forall \delta \in [0,1]$

We can also derive the *additive* form of the bound by setting $t = \delta\mu$:

- **Additive Chernoff bound**: $Pr[|X - \mu| \geq t] \leq 2\exp\left(-\frac{t^2}{3\mu}\right), \quad \forall t \in [0,\mu]$

Union Bound

Let $E_1, \ldots, E_n$ be arbitrary events in some probability space. Then,

$$Pr\left[\bigcup_{i=1}^{n} E_i\right] \leq \sum_{i=1}^{n} Pr[E_i]$$

That is, for a set of events $E_1, \ldots, E_n$, the probability of at least one event occurring is less than or equal to the sum of the probabilities of the individual events. We can think of the union bound with a simple set-based analogy i.e. if events are viewe as subsets of a potential sample space, then the union of these sets can be no larger than the sum of the individual sets (i.e. since the sets may have some non-empty intersections).

References

[1] Benjamin Doerr. Probabilistic tools for the analysis of randomized optimization heuristics. In *Natural Computing Series*, pages 1–87. Springer International Publishing, nov 2019.