0.1 Expectation Values of Functions of Random Variables

If we have a function $F(A)$ on a random variable,

$$\langle F \rangle = \sum_f f \Pr_F(f)$$

$$= \sum_f f \sum_a \delta_{f,F(a)} \Pr_A(a)$$

$$= \sum_a \Pr_A(a) \underbrace{\sum_f f \delta_{f,F(a)}}_{F(a)}$$

$$= \sum_a F(a) \Pr_A(a)$$

Therefore, we can calculate the expectation value of F using either the probability distribution of the function itself or by using the probability distribution of the random variable.

0.1.1 Moments

The n-th moment is defined as

$$\langle F^n \rangle = \sum_f f^n \Pr_F(f) = \sum_a [F(a)]^n \Pr_A(a)$$

We can then define "centered" moments as

$$\langle (F - \langle F \rangle)^n \rangle = \cdots$$

The most important of these is the variance, the second centered moment:

$$\sigma_F^2 = \left\langle (F - \langle F \rangle)^2 \right\rangle = \left\langle F^2 - 2F \langle F \rangle + \langle F \rangle^2 \right\rangle$$

Notice that $\langle F \rangle^2$ is just a number, so we already know its expectation value. Again, recall that expectation values are linear so we can also remove the $2 \langle F \rangle$ factor:

$$\sigma_F^2 = \left\langle F^2 \right\rangle - \langle F \rangle^2$$

From the original definition, it is obvious that the variance is strictly non-negative. If the variance vanishes, the random variable never deviates from its expectation value, so it is not random. We can define the square root of the variance as the standard deviation of F: σ_F. It's slightly nicer than the variance because it has the same units as the random variable, but often the square root makes it difficult to work with.

Definition 0.1.1 (Covariance)**.**

$$\text{Cov}(A, B) = \langle AB \rangle - \langle A \rangle \langle B \rangle$$

For a single variable, $\sigma_A^2 = \text{Cov}(A, A)$.

> **Definition 0.1.2** (Correlation Coefficient)**.**
> $$c_{A,B} = \frac{\text{Cov}(A,B)}{\sigma_A \sigma_B}$$

This coefficient is dimensionless and invariant under linear transformations of the random variables:

$$A \mapsto a_0 + a_1 A \quad \text{and} \quad B \mapsto b_0 + b_1 B \implies c_{A,B} \mapsto c_{A,B}$$

Additionally,

$$-1 \leq c_{A,B} \leq +1$$

> **Definition 0.1.3** (Correlation)**.** If the covariance vanishes, we say that the variables are uncorrelated.
>
> Independence implies non-correlation, but non-correlation does not imply independence.
>
> The proof of the forward direction is a result of the joint probabilities factoring to allow the covariance to cancel to zero. The disproof of the opposite direction can be made by clever counterexample. Correlation is a measure of linear dependence, but two random variables can be non-linearly dependent.

0.1.2 Sums of Random Variables

$$S = \sum_{j=1}^{N} F_j$$

$$\langle S \rangle = \sum_{j=1}^{N} \langle F_j \rangle$$

$$\sigma_S^2 = \langle S^2 \rangle - \langle S \rangle^2$$
$$= \left\langle \sum_{j=1}^{N} F_j \sum_{k=1}^{N} F_k \right\rangle - \left\langle \sum_{j=1}^{N} F_j \right\rangle \left\langle \sum_{k=1}^{N} F_k \right\rangle = \sum_{j,k} \left(\langle F_j F_k \rangle - \langle F_j \rangle \langle F_k \rangle \right) = \sum_{j,k} \text{Cov}(F_j, F_k)$$

If all $\{F_j\}$ are uncorrelated, then $\text{Cov}(F_j, F_k) = \delta_{jk} \sigma_{F_j}^2$. In this special case,

$$\sigma_S^2 = \sum_{j,k} \delta_{jk} \sigma_{F_j}^2 = \sum_j \sigma_{F_j}^2$$

Let's now look at the special case where all of the random variables is independent and each has an identical probability distribution. For example, if we roll a bunch of identical dice, the results are all statistically independent, but the probabilities are the same.

$$\langle F_j \rangle = \mu$$

Define

$$\sigma_{F_j}^2 = \sigma^2$$
$$\langle S \rangle = N\mu$$

so

$$\sigma_S^2 = N\sigma^2 \implies \sigma_S = \sqrt{N}\sigma$$

This also implies that

$$\frac{\sigma_S}{\langle S \rangle} = \frac{1}{\sqrt{N}} \frac{\sigma}{\mu}$$

Let's look at an even more special case, where each random variable has a binomial distribution. In this case, each F_j corresponds to N individual success/fail experiments where p is the probability of success. In other words, there are two possible values with probabilities p and $1 - p$. Call the success outcome 1 and failure 0 such that $\Pr(1) = p$ and $\Pr(0) = 1 - p$.

$$S = \sum_{j=1}^{N} = F_j$$

so

$$\langle S \rangle = N \langle F_j \rangle = N(1 \cdot p + 0 \cdot (1 - p)) = Np$$

$$\sigma_S^2 = N\sigma_{F_j}^2 = N \cdot \left(\langle F_j^2 \rangle - \langle F_j \rangle^2 \right) = Np(1 - p)$$

so

$$\frac{\sigma_S}{\langle S \rangle} = \frac{1}{\sqrt{N}} \cdot \sqrt{\frac{1 - p}{p}}$$

As N gets bigger, the sum gets less random!

We can work out the exact probability distribution:

$$\Pr(S, N; p) = \binom{N}{S} p^S (1 - p)^{N-S} \qquad \text{(Binomial Distribution)}$$

where

$$\binom{N}{S} = \frac{N!}{S!(N - S)!}$$

Let's make sure this distribution is normalized. We can do this using the binomial theory:

$$(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

so

$$\sum_{S=0}^{N} \Pr(S, N; p) = \sum_{S=0}^{N} \binom{N}{S} p^S (1 - p)^{N-S} = [p + (1 - p)]^n = 1$$

Definition 0.1.4 (Stirling's Approximation for Factorials)**.**

$$\ln N! = \ln (1 \cdot 2 \cdot 3 \cdot \dots \cdot N)$$
$$= \ln 1 + \ln 2 + \dots + \ln N$$
$$= \sum_{n=1}^{N} \ln n$$
$$\approx \int_{1}^{N} \mathrm{d}n \ln n = N \ln N - N + 1 \approx N \ln N - N$$

if N is large.