1 CONDITIONAL PROBABILITY

Definition 1.0.1 For any two events, A & B, for a random variable, W, defined on a sample space, Ω, the conditional probability of A given B is $P(W \in A \mid W \in B) \overset{\text{defn}}{=} \dfrac{P(W \in AB)}{P(W \in B)}$, provided $P(W \in B) \neq 0$

Often abbreviated as $P(A|B) = P(AB)/P(B)$, provided $P(B) \neq 0$.

A trivial algebraic calculation shows $P(AB) = P(A|B)P(B) = P(B|A)P(A)$, provided $P(A)P(B) > 0$

Theorem 1.0.1 The Bayes Theorem

Given any partition of the sample space Ω into non-trivial mutually disjoint events B_n, $n \in \mathbb{N}$, then for any other event $A \subset \Omega$ we obtain

1) total probability $P(A) = \sum_{n \in \mathbb{N}} P(A|B_n)P(B_n)$

   Proof: $A = A\Omega = A\sum_1^\infty B_n = \sum_1^\infty AB_n \Rightarrow P(A) = \sum_1^\infty P(AB_n) = \sum_1^\infty P(A|B_n)P(B_n)$. $\square$

2) bayes formula $P(B_n|A) = P(A|B_n)P(B_n)/P(A)$

If we view event A as a red light that goes on w/ a certain probability @ any B_n, $n \in \mathbb{N}$ of a certain process happens to fail, then the certainty $P(B_n|A)$ tells us how often B_n would be the culprit to the breakdown.

A is said to be observable, while the non-observable B_n, $n \in \mathbb{N}$ is to be inferred.

2 JOINT DISTRIBUTIONS & STATISICAL INDEPENDENCE

Consider the circumstance that the given variable is an ordered pair, $W = (X, Y)$.

W is said to enjoy a bivariate/joint distribution with coordinates/components X & Y.

We can imagine each of these coordinate variables to be distributed on its own individual sample space, $\Omega = \mathcal{X} \times \mathcal{Y}$

The individual distributions are referred to as marginal distributions in respect of the joint distribution from which they derive. Given that $W = (X, Y) \sim P$ on $\Omega = \mathcal{X} \times \mathcal{Y}$, we automatically have marginal distributions

$$X \sim P_X \text{ on } \mathcal{X},\quad P_X(A) = P(X \in A) = P((X, Y) \in A \times \mathcal{Y}) \quad \forall A \subset \mathcal{X}$$

$$Y \sim P_Y \text{ on } \mathcal{Y},\quad P_Y(B) = P(Y \in B) = P((X, Y) \in \mathcal{X} \times B) \quad \forall B \subset \mathcal{Y}$$

We say variable Y has no stochastic influence on variable X, "X is statistically independent of Y," iff for every event B for Y, for which $P(Y \in B) \neq 0$, the condidional probabilities/frequencies for X given $Y \in B$ are the same as its original marginal probabilities/frequencies.

Definition 2.0.1 X is said to be statistically independent of Y, denoted $X \perp\!\!\!\perp Y$, if and only if

$$P(X \in A \mid Y \in B) = P(X \in A), \quad \forall A \subset \mathcal{X}, B \subset \mathcal{Y} \text{ w. } P(Y \in B) \neq 0$$

Proposition 2.0.1 (multiplicative criterion of independence)

$$X \perp\!\!\!\perp Y \iff P(X \in A, Y \in B) = P(X \in A)P(Y \in B) \quad \forall A \subset \mathcal{X}, B \subset \mathcal{Y}$$

Through symmetry, $X \perp\!\!\!\perp Y \iff Y \perp\!\!\!\perp X$. It follows that if $P(Y \in B) \neq 0$ & $P(Y \in B^c) \neq 0$:

$$\begin{aligned} P(X \in A \mid Y \in B) = P(X \in A) &\iff P(X \in A \mid Y \in B^c) = P(X \in A) \\ &\iff P(X \in A^c \mid Y \in B^c) = P(X \in A^c) \\ &\iff P(X \in A^c \mid Y \in B^c) = P(X \in A^c) \end{aligned}$$

Proposition 2.0.2 (atomic characterization of independence)

$$X \perp\!\!\!\perp Y \iff I_A(X) \perp\!\!\!\perp I_B(Y), \quad \forall A \subset \mathcal{X}, B \subset \mathcal{Y}$$

Proposition 2.0.3 (invariance of statistical independence)

$$X \perp\!\!\!\perp Y \implies \phi(X) \perp\!\!\!\perp \psi(Y), \ \forall \phi: \mathcal{X} \to \mathcal{U}, \ \psi: \mathcal{Y} \to \mathcal{V}$$

If X & Y are both real-valued, then from multiplicative criteron, statistical independence implies a factorization of the joint distribution function, $F(x,y) \overset{\text{defn}}{=} P(X \le x, Y \le y)$, $(x,y) \in \mathbb{R}^2$, as a product of the two individual marginal DFs.

$$X \perp\!\!\!\perp Y \implies F(x,y) = F(x)F(y), \ \forall (x,y) \in \mathbb{R}^2.$$

Proposition 2.0.4 (statistical independence + product expectations)

$$F(x,y) = F_X(x)F_Y(y) \ \forall (x,y) \in \mathbb{R}^2 \iff \mathbb{E}g(X)h(Y) = \mathbb{E}g(X)\mathbb{E}h(Y), \ \forall g: \mathbb{R} \to \mathbb{R}, \ h: \mathbb{R} \to \mathbb{R}$$

Corallary 2.0.4.1 (statistical independence + product expectations)

$$X \perp\!\!\!\perp Y \iff \mathbb{E}g(X)h(Y) = \mathbb{E}g(X)\mathbb{E}h(Y), \quad \forall g: \mathcal{X} \to \mathbb{R}, \ h: \mathcal{Y} \to \mathbb{R}$$

MULTIVARIATE INDEPEDENCE

Consider the mutual statistical independence of arbitrarily indexed collections of RV, $X = (X_t, t \in T)$.

Each component, X_t, is on its own individual sample space, $\mathcal{X}_t$, for each t in some index set T

The object/T-sequence $X = (X_t, t \in T)$ is said to have a joint distribution on the product space.

Each component, X_t, for each $t \in T$, has its separate marginal distribution on its own individual component space.

For any particular finite subset of the index set, $S \subset T$, $S = \{s_1, \dots, s_n\}$, consider the subordinate variable

$$X_S = (X_s, s \in S) = (X_{s_1}, \dots, X_{s_n})$$

The mutual statistical independence of all components of original $X = (X_t, t \in T)$ is simply the basis bivariate independence of all possible pairs of non-intersecting finite subordinate variables that may be formed of it.

Definition 2.1.1 $X = (X_t, t \in T)$ is said to be mutually statistically independent, $(X_t, t \in T)$ indep,

$$\text{iff} \quad X_{S_1} \perp\!\!\!\perp X_{S_2} \quad \forall S_1, S_2 \overset{\text{finite}}{\subset} T, \ \text{s.t. } S_1 \cap S_2 = \emptyset$$

Definition 2.1.2 A sequence of RV $X = (X_t, t \in T)$ is said to be identically + independently distributed (iid) w/ common distribution identical to that of some specific other X iff:

i) identically distributed: $X_t \overset{d}{=} X$ for every $t \in T$

ii) mutually independent: $(X_t, t \in T)$ independent

} denoted $X_t, t \in T$ iid X

One may show that the finite ~~or infinite~~ sequence $X_1, \dots, X_n$ will be mutually statistically independent iff

$$\mathbb{E}g_1(X_1)\cdots g_n(X_n) = \mathbb{E}g_1(X_1)\cdots\mathbb{E}g_n(X_n) \quad \text{for any real-valued func. } g_1, \dots, g_n$$

Notice $(X_n, n \in \mathbb{N})$ indep $\iff (X_1, \dots, X_n) \perp\!\!\!\perp X_{n+1} \ \forall n \in \mathbb{N}$.

In the particular case where $X_1, \dots, X_n$ iid X, the finite sequence may be referred to variously as the generic distribution, common distribution, etc. Notice the general consequence

For any given events $A_1, \dots, A_n, \dots$ on the generic sample sample $\mathcal{X}$, we have the conjunction

$$(X_1 \in A_1) \cap \cdots \cap (X_n \in A_n) = ((X_1, \dots, X_n) \in A_1 \times \cdots \times A_n)$$

Thus, if $X_1, \dots, X_n, \dots$ iid X, then

$$\begin{aligned} P((X_1, \dots, X_n) \in (A_1, \dots, A_n)) &= P((X_1 \in A_1) \cap \cdots \cap (X_n \in A_n)) \\ &= P(X_1 \in A_1) \cdots P(X_n \in A_n) \\ &= P(X \in A_1) \cdots P(X \in A_n) \end{aligned}$$

More generally, $\forall g_1, \dots, g_n: \mathcal{X} \to \mathbb{R}$ real-valued on $\mathcal{X}$, $\mathbb{E}g_1(X_1)\cdots g_n(X_n) = \prod_{i=1}^{n} \mathbb{E}g_i(X_i)$.

All calculations of specific probabilities/expected values for multivariate $(X_1, \dots, X_n)$ are ultimated to be obtained in terms of the single generic univariate X itself.

3 Bionomial Distribution

Consider a random variable S_n that counts the number of successes in n iid bernoulli trials: $z_i, i \in \mathbb{N}$ iid $z \sim \text{bern}(p)$.

You can easily see that $S_n = z_1 + \cdots + z_n$ is the object whose random output gives the # of 1's that occur for vector $z = (z_1, \ldots, z_n)$ for the first n of these bernoulli trials

S_n is the number of success in n trials

Definition 3.0.1 The random variable X is said to have a binomial distribution on n trials, w/ probability of success per trial, p, — denoted $X \sim \text{bin}(n, p)$, $n \in \mathbb{N}$, $0 \le p \le 1$ iff

$$X \overset{d}{=} S_n = z_1 + \cdots + z_n \quad \text{w.} \quad z_1, \ldots, z_n \text{ iid } z \sim \text{bern}(p)$$

It immediately follows that $\mathbb{E}X = \mathbb{E}S_n = \mathbb{E}(z_1 + \cdots + z_n) = \mathbb{E}z_1 + \cdots + \mathbb{E}z_n = n\mathbb{E}z = np$

The long-run average number of successes is the average number of successes per trial, p, times # of trials, n

Intuitively, the sum of two independent binomials that have the same p is still binomial

Proposition 3.0.1 (sums of independent binomials)

$$X \sim \text{bin}(m, p),\ Y \sim \text{bin}(n, p),\ X \perp\!\!\!\perp Y \implies X + Y \sim \text{bin}(m+n, p)$$

Proof Take any $(z_i, i \in \mathbb{N})$ iid $z \sim \text{bern}(p)$ so that

$$\begin{pmatrix} X \\ Y \end{pmatrix} \overset{d}{=} \begin{pmatrix} z_1 + \cdots + z_m \\ z_{m+1} + \cdots + z_n \end{pmatrix} \qquad \text{then } X + Y \overset{d}{=} z_1 + \cdots + z_{m+n}$$

Corollary 3.0.1.1 $X_i \overset{\text{indep}}{\sim} \text{bin}(n_i, p),\ i = 1, \ldots, k \implies \sum_{i=1}^{k} X_i \sim \text{bin}\left(\sum_{i=1}^{k} n_i, p\right)$

Any single outcom of n-vectors of z-trials will be one of the 2^n potential sequences of 0's + 1's in $\mathbb{R}^n$.

$\therefore$ the sample space for RV $z = (z_1, \ldots, z_n)$ consists of 2^n corners, $0 \times 1 = \{0,1\}^n = 2^n$ of unit cube $[0,1]^n \in \mathbb{R}^n$

The trials are independent, so the frequency that any particular corners occurs is:

$$P(z = z) = \prod_{i=1}^{n} P(z_i = z_i) \quad \forall z \in 0 \times 1, \quad \text{individually} \quad P(z_i = z_i) = p^{z_i} q^{1-z_i} \quad i = 1, \ldots, n$$

Thus the full probability mass function of vector variable z over all 2^n elements of its domain is

$$P(z = z) = p^{\Sigma z_i} q^{n - \Sigma z_i} \qquad \forall z \in 0 \times 1$$

Now we are enabled to determine the pmf for the counting variable $S_n = z_1 + \cdots + z_n$.

Note that S_n will have the value $k \iff z$ lands on any one of the specific corners of unit cube w/ k 1's + $n-k$ 0's.

Accordingly, let $C_k^n = \{z \in 2^n \mid \Sigma z_i = k\}$ for each $k = 0, 1, \ldots, n$. We begin to compute $S_n = k \iff z \in C_k^n$

$$\therefore \quad P(S_n = k) = P(z \in C_k^n) = \sum_{z \in C_k^n} P(z = z) = \sum_{\{z \mid \Sigma z_i = k\}} p^{\Sigma z_i} q^{n \Sigma z_i} = {}^{\#}C_k^n \cdot p^k q^{n-k}$$

Combinations (cardinalities)

Previously known as binomial coefficients, pascal #'s, triangle #'s, etc, the cardinalities ${}^{\#}C_k^n$, $k = 0, 1, \ldots, n$ are often denoted as

$$\binom{n}{k} \overset{\text{name}}{=} {}^{\#}C_k^n$$

read as "n choose k"

$\forall n \in \mathbb{N}$, $\forall 0 \le p \le 1$, we have $\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} P(S_n = k) = 1$. On a simple distribution $p = \frac{t}{1+t}$, $t \ge 0$, we get $(1+t)^n = \sum_{k=0}^{n} \binom{n}{k} t^k$.

Taking derivatives on both sides, we get

$$\binom{n}{k} = \frac{n}{k} \binom{n-1}{k-1} = \frac{n^{(k)}}{k!}$$

Descending Factorial

If $X \sim \text{bin}(n, p)$, then $\forall r = 1, 2, 3, \ldots$, the rth descending/falling factorial is $X^{(r)} = X(X-1)\cdots(X-r+1)$

$$\mathbb{E}X^{(r)} = \sum_{k=0}^{n} k^{(r)} \binom{n}{k} p^k q^{n-k} = \cdots = n^{(r)} p^r \sum_{k=r}^{n} \frac{(n-r)^{(k-r)}}{(k-r)!} p^{k-r} q^{n-k} = n^{(r)} p^r \ \text{ for } r \le n, \quad \mathbb{E}X^{(r)} = 0 \text{ for } r > n$$

Notice $X^{(2)} = X^2 - X$, $n^{(2)} p = \mathbb{E}X^{(2)} = \mathbb{E}X^2 - \mathbb{E}X = \mathbb{E}X^2 - np \implies \mathbb{E}X^2 = n^{(2)} p^2 + np$

4 GEOMTRIC + NEGATIVE-BIONOMIAL DISTRIBUTIONS (waiting times w/ bernoulli trials)

Given $(Z_n, n \in \mathbb{N})$ iid $Z \sim$ bernoulli(p), for each $k \in \mathbb{N}$, let T_k count the number of random trials required until + including the k^{th} success.

$$T_1 = \min\{n \in \mathbb{N}, Z_n = 1\}$$

$$T_k = \min[\{n \in \mathbb{N}, Z_n = 1\} - \{T_1, \dots, T_{k-1}\}] \quad k \geq 2$$

Definition 4.0.1 The random variable Y is said to have negative-binomial distribution on k successes,

— denoted $Y \sim \text{negbin}(k, p)$, $0 < p \leq 1$, $k \in \mathbb{N}$ iff $Y \overset{d}{=} T_k$, defined by

$$T_k = n \equiv \sum_{i=1}^{n-1} Z_i = k-1, \; Z_n = 1 \qquad \forall n \geq k$$

The probability mass function for $Y \sim \text{negbin}(k,p)$, $0 < p \leq 1$, $k \in \mathbb{N}$ is defined by

$$\Leftrightarrow P(Y=n) = P(T_k = n) = P\left(\textstyle\sum_1^{n-1} Z_i = k-1, Z_n = 1\right) = \binom{n-1}{p-1} p^k q^{n-k}, \qquad n = k, k+1, \dots$$

probability of the n^{th} trial being the k^{th} succes

k-1 successes in the first n-1 trials

$$\binom{n-1}{p-1} p^{k-1} q^{n-1-(k-1)} \quad p$$

last trial being a success

k-1 success can be in any order

When $k=1$, $P(T_1 = n) = P\left(\sum_{i=1}^{n-1} Z_i = 0, Z_n = 1\right) = pq^{n-1}$, $n = 1, 2, \dots$

The only success in the n trials must be the last one, so only one possible order

Definition 4.0.2 The random variable W is said to have a geometric distribution — denoted $W \sim \text{geo}(p)$, $0 < p < 1$

iff $W \overset{d}{=} T_1$

W is stochastically identical to T_1, so $\text{geo}(p) \equiv \text{negbin}(1, p)$, the # of random trials until the first success.

$$\mathbb{E}W = \sum_{i=1}^{\infty} i p q^{i-1} = p \sum_{i=1}^{\infty} i q^{i-1} = p \frac{d}{dq} \sum_{i=1}^{\infty} q^i = p \frac{d}{dq} \frac{1}{1-q} = \frac{1}{p}$$

Proposition 4.0.1 (sum of independent negative-binomials 1)

$$T \sim \text{negbin}(k,p),\ W \sim \text{geo}(p),\ T \perp\!\!\!\perp W \Rightarrow T + W \sim \text{negbin}(k+1, p)$$

Proof: Take any two independent bern(p) sequences: $(Z_i, i \in \mathbb{N}) \perp\!\!\!\perp (Z_i', i \in \mathbb{N})$. Then for each $n = k+1, \dots$

$$\begin{aligned} P(T+W, n) &= \sum_{i=1}^{\infty} P(T+W=n, W=i) = \sum_{i=1}^{\infty} P(T=n-i) P(W=i) \\ &= \sum_{j=k}^{n-1} P(T=j) P(W=n-j) \\ &= \sum_{j=k}^{n-1} P(Z_1 + \dots + Z_{j-1} = k-1, Z_j = 1) P(Z_1' + \dots + Z_{n-j-1}' = 0, Z_{n-j}' = 1) \\ &= \sum_{j=k}^{n-1} P(S_{j-1} = k-1, Z_j = 1, Z_{j+1} + \dots + Z_{n-1} = 0, Z_n = 1) \\ &= P(S_{n-1} = k, Z_n = 1) = P(T_{k+1} = n) \end{aligned}$$

Corollary 4.0.1.1 (structure of the negative binomial)

$$T \sim \text{negbin}(k,p),\ 0 < p \leq 1,\ k \in \mathbb{N} \Leftrightarrow T \overset{d}{=} \sum_{i=1}^{k} W_i \quad \text{w. } W_i \text{ iid } W \sim \text{geo}(p)$$

Corollary 4.0.1.2 (sum of independent negative-binomials 2)

$$S \sim \text{negbin}(k,p),\ T \sim \text{negbin}(\ell, p),\ S \perp\!\!\!\perp T \Rightarrow S + T \sim \text{negbin}(k+\ell, p)$$

Corollary 4.0.1.3 (sums of independent negative-binomials 3)

$$T_i \overset{\text{indep}}{\sim} \text{negbin}(k_i, p) \Rightarrow \sum_{i=1}^{n} T_i \sim \text{negbin}\left(\textstyle\sum k_i, p\right)$$

Through $\mathbb{E}W$ and $\mathbb{E}Z$, we know p = average # success per trial, $\frac{1}{p}$ = average # trials per success.

Let $f(q) = \frac{1}{1-q} = \sum_{i=0}^{\infty} q^i \quad |q| < 1$. For $r = 0, 1, 2, \dots$

(not derivative!) $f^{(r)}(q) = r!/(1-q)^{r+1} = r! f(q)^{r+1}$, $|q| < 1$

$$= \sum_{i=r}^{\infty} i^{(r)} q^{i-r} = \sum_{i=r}^{\infty} \frac{i!}{(i-r)!} q^{i-r}$$

We immediately acquire $\mathbb{E}W^{(r)} = \sum_{i=1}^{\infty} i^{(r)} p q^{i-1} = \dots = r! \frac{q^{r-1}}{q^r}$

In summary: $\mathbb{E}W = \frac{1}{p}$, $\mathbb{E}W(W-1) = \frac{2p}{p^2}$, $\mathbb{E}W^2 = \frac{1+q}{p^2}$, $\mathbb{E}T_k = k\mathbb{E}W$

There is a more immediate connection between the negative binomial family $(T_k, k \in \mathbb{N})$, & the binomial family $(S_n, n \in \mathbb{N})$

We wait longer than n trials to get k successes $(T_k > n)$ $\iff$ the number within n trials is smaller than k $(S_n < k)$

Thus, we have the fundamental conjugate relation between the two variables

$$T_k > n \iff S_n < k \qquad \forall k \in \mathbb{N}, n \in \mathbb{N}$$

there are more than n trials until the k^{th} success $\iff$ there are less than k successes in n trials

Proposition 4.0.2 (binomial / negative-binomial connection)

For discrete conjugal processes $\quad Y_k > n \iff X_n < k \quad \forall k \in \mathbb{N}, n \in \mathbb{N}$, we have the specific connection in distribution

$$X_n \sim \text{binomial}(n, p) \;\; \forall n \in \mathbb{N} \iff Y_k \sim \text{negbin}(k, p) \;\; \forall k \in \mathbb{N}$$

(S_n) X_n = random number of successes in fixed n trials

(T_k) Y_k = random number of trials for fixed k successes

Proof $(\Rightarrow)$

$$\begin{aligned} P(Y_k = n) &= P(Y_k \le n) - P(Y_k \le n-1) \qquad n = k, k+1, \dots \\ &= P(Y_k > n-1) - P(Y_k > n) \\ &= P(X_{n-1} < k) - P(X_n < k) \\ &= P(S_{n-1} < k) - P(S_n < k) \\ &= P(S_{n-1} < k \le S_n) = P(S_{n-1} = k-1, Z_n = 1) \qquad \blacksquare \end{aligned}$$

$(\Leftarrow)$

$$\begin{aligned} P(X_n = k) &= P(X_n < k+1) - P(X_n < k) \\ &= P(Y_{k+1} > n) - P(Y_k > n) = P(T_{k+1} > n) - P(T_k > n) \\ &= P(T_k \le n) - P(T_{k+1} \le n) \\ &= P(T_k \le n-1) - P(T_{k+1} \le n-1) + P(T_k = n) - P(T_{k+1} = n) \\ &= P(X_{n-1} = k) + P(T_k = n) - P(T_{k+1} = n) \\ &\overset{\text{ind}}{=} P(S_{n-1} = k) + P(S_{n-1} = k-1, Z_n = 1) - P(S_{n-1} = k, Z_n = 1) \\ &= P(S_{n-1} = k, Z_n = 0) + P(S_{n-1} = k-1, Z_n = 1) = P(S_n = k) \qquad \blacksquare \end{aligned}$$

5 Binomial Process

Given a sequence $(Z_n, n\in\mathbb{N})$ iid $Z\sim \text{bern}(p)$ $0<p<1$, consider the derived sequence of partial sums $S_n = Z_1+\cdots+Z_n$, $n\in\mathbb{N}$.
Letting $S_0=0$, recall

$$X\sim \text{bin}(n,p) \iff P(X=k)=P(S_n=k) \quad \forall k\in\mathbb{N}$$

$$\begin{aligned} T\sim \text{negbin}(k,p) \iff P(T=n) &= P(S_{n-1}=k-1, Z_n=1) \quad \forall n\in\mathbb{N} \\ &= P(S_{n-1}=k-1, S_n=k) \\ &= P(S_{n-1}<k\le S_n) \end{aligned}$$

Definition 5.0.1 A binomial process is any sequence of random variables whose joint distribution is identical to the joint distributions of the S_n's

$$(X_n, n\in\mathbb{N}) \overset{d}{=} (S_n, n\in\mathbb{N})$$

If we take any subsequence $(X_{n_i}, i\in\mathbb{N})$, then $(X_{n_i}, i\in\mathbb{N}) \overset{d}{=} (S_{n_i}, i\in\mathbb{N})$. Therefore, in turn (setting $n_0=0, X_0=0, T_0=0$),

$$(X_{n_{i+1}}-X_{n_i}, i\in\mathbb{N}) \overset{d}{=} (S_{n_{i+1}}-S_{n_i}, i\in W) = \left(\sum_{j=n_i+1}^{n_{i+1}} Z_j, i\in W\right)$$

The incremental differences, $X_{n_{i+1}}-X_{n_i}$, $i\in W$, are mutually, statistically independent binomials, may be so denoted

$$X_{n_{i+1}}-X_{n_i} \overset{\text{indep}}{\sim} \text{bin}(n_{i+1}-n_i, p) \quad i\in W$$

⊙ we take any subsequence $(Y_{k_j}, j\in\mathbb{N})$ from the waiting time process, $(Y_k, k\in\mathbb{N})$, conjugate to some given binomial process, the successive epochs, $Y_{k_{j+1}}-Y_{k_j}$, $j\in\mathbb{N}$, prove to be mutually statistically independent negative binomials.

Theorem 5.0.1 (reciprocity in the binomial/neg-binomial process)

For reciprocal process $Y_k>n \iff X_n<k$, $\forall k\in\mathbb{N}, n\in\mathbb{N}$, we have the specific connection in distribution

$$X_{n_{i+1}}-X_{n_i} \overset{\text{indep}}{\sim} \text{bin}(n_{i+1}-n_i, p) \quad \forall\, 0<n_i\uparrow\uparrow \quad \text{(increasing sequence)}$$

$$\iff$$

$$Y_{k_{j+1}}-Y_{k_j} \overset{\text{indep}}{\sim} \text{negbin}(k_{j+1}-k_j, p) \quad \forall\, 0<k_j\uparrow\uparrow$$

Proof on pg 31

Problem of the Points

Suppose A + B are playing a game, victory for A requires A wins M points before B wins N. Let the entire game be regarded as a sequence of iid bernoulli trials, Z_i iid $Z\sim\text{bern}(p)$, $Z_i=A$ is a success for A. Let T_M be # trials before A has M wins.

$$\text{victory for } A \iff T_M=M \text{ or } T_M=M+1 \text{ or} \dots \text{ or } T_M=M+N+1$$

$$\text{overall} \quad P(\text{victory for } A) = P(M\le T_M\le M+N-1) = P(S_{M+N-1}\ge M)$$

Summary

$$W\sim\text{geo}(p) \iff P(W>n)=P(S_n<1)=P(S_n=0)=q^n$$

$$\iff P(W\ge n)=q^{n-1}$$

$$\text{(exercise)} \iff P(W\ge n) = p\sum_{k=n}^{\infty} P(W\ge k)$$

$$\mathbb{E}W = \sum_{n=1}^{\infty} P(W\ge n) = \frac{P(W\ge 1)}{p} = \frac{1}{p}$$

$$\begin{aligned} \mathbb{E}W^2 &= \sum_{n=1}^{\infty} n^2 P(W=n) = \sum_{n=1}^{\infty}\sum_{m=1}^{n}(2m-1)P(W=n) \\ &= 2\sum_{m=1}^{\infty} mP(W\ge m) - \sum_{m=1}^{\infty}P(W\ge m) \\ &= 2\sum_{m=1}^{\infty}\sum_{k=1}^{m} P(W\ge m) - \frac{1}{p} \\ &= 2\sum_{k=1}^{\infty}\sum_{m=k}^{\infty} P(W\ge m) - \frac{1}{p} \\ &= \frac{2}{p}\sum_{k=1}^{\infty} P(W\ge k) - \frac{1}{p} = \frac{2}{p}\frac{1}{p} - \frac{1}{p} \\ &= \frac{2-p}{p^2} = \frac{1+q}{p^2} \end{aligned}$$

6 Poisson Distribution

Suppose for a specified unit, we throw an enormous number n of iid bernoull trials at a target and find that the average number of successes, λ, is relatively very small.

Let X denote the random number of successes that occur within the chosen unit. X has binomial distribution w/ sample size n. Let p denote the probability of success per trial. Then, it follows that

$$\mathbb{E}X = np = \lambda \qquad p = \frac{\lambda}{n}$$

Any attempt to evalute various probabilities of X must depend on parameter $\lambda \propto \frac{1}{n}$.

Consider the explicit computation of individual probabilities @ n is "huge" + $\lambda = np$ is much smaller

$$\begin{aligned} P(X=k) &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= \cdots \\ &= \left(1-\frac{\lambda}{n}\right)^n \frac{\lambda^k}{k!} \frac{(1-\frac{1}{n})\cdots(1-\frac{k-1}{n})}{(1-\frac{\lambda}{n})\cdots(1-\frac{\lambda}{n})} \\ &\approx e^{-\lambda}\frac{\lambda^k}{k!} \end{aligned}$$

Definition 6.0.1 The random variable N is said to have poisson distribution w/ average number of successes $\lambda \geq 0$ — denoted $N \sim \text{poisson}(\lambda)$, $\lambda \geq 0$, iff

$$P(N=k) = e^{-\lambda}\frac{\lambda^k}{k!} \qquad k = 0, 1, 2, \ldots$$

This gives us the fundamental characterization

$$N \sim \text{poisson}(\lambda) \iff P(N=k) = \lim_{n\to\infty} P(X_n = k)$$

$$\text{w. } X_n \sim \text{bin}\left(n, \frac{\lambda}{n}\right), \ n \in \mathbb{N}$$

$\mathbb{E}$ of poisson may be computed as they were for bin/negbin — in rth descending factorial

$$N^{(r)} = N(N-1)\cdots(N-r+1), \ r \in \mathbb{N}$$

$$\begin{aligned} \mathbb{E}N^{(r)} &= \sum_{k=0}^{\infty} k^{(r)} e^{-\lambda}\frac{\lambda^k}{k!} \\ &= \sum_{k=r}^{\infty} k^{(r)} e^{-\lambda}\frac{\lambda^k}{k!} \\ &= \lambda^r \sum_{k=r}^{\infty} e^{-\lambda}\frac{\lambda^{k-r}}{(k-r)!} = \lambda^r \end{aligned}$$

$$\implies \lambda^2 = \mathbb{E}N^{(r)} = \mathbb{E}N^2 - \mathbb{E}N = \mathbb{E}N^2 - \lambda$$

$$\mathbb{E}\lambda^2 = \lambda^2 + \lambda$$

Proposition 6.0.1 (sums of independent poissons)

$$\left.\begin{aligned} &N_1 \sim \text{poisson}(\lambda_1) \\ &N_2 \sim \text{poisson}(\lambda_2) \\ &N_1 \perp\!\!\!\perp N_2 \end{aligned}\right\} \implies \begin{aligned} &\text{i) } N_1 + N_2 \sim \text{poisson}(\lambda_1+\lambda_2) \\ &\text{ii) } N_i \mid (N_1+N_2 = n) \sim \text{binomial}\left(n, \frac{\lambda_i}{\lambda_1+\lambda_2}\right) \end{aligned}$$

Proof

i) $$\begin{aligned} P(N_1+N_2=n) &= \sum_{k=0}^{\infty} P(N_1+N_2=n, N_1=k) = \sum_{k=0}^{n} P(N_1=k)P(N_2=n-k) = \sum_{k=0}^{n} e^{-\lambda_1}\frac{\lambda_1^k}{(n-k)!} e^{-\lambda_2}\frac{\lambda_2^{n-k}}{k!} \\ &= e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!}\sum_{k=0}^{n}\binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k} \\ &= e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!} \end{aligned}$$

ii) $$P(N_1=k \mid N_1+N_2=n) = \frac{P(N_1=k, N_1+N_2=n)}{P(N_1+N_2=n)} = \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$$

POISSON PROCESS

Let N_t denote the number of successes in any interval of "length" $t \geq 0$. Suppose that the expected value of N_t is directly proportional to that length, w/ constant of proportionality $\lambda \geq 0$. So, $EN_t = \lambda t \quad \forall t \geq 0$

If N_t arises as the sum of an enormous # of independent bernoulli trials, for which probability of individual successes on any trial is diminishingly small, then we can consider N_t to be poisson distributed

$$N_t \sim \text{poisson}(\lambda t), \quad t \geq 0, \ \lambda \geq 0$$

A poisson process in $t \geq 0$ is completely described by adding only that

i) $\forall\, 0 \leq s \leq t: \quad N_t - N_s \stackrel{d}{=} N_{t-s}$

ii) for $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$, $\quad N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_{t_n} - N_{t_{n-1}}$ indep

7 GAMMA & BETA DISTRIBUTIONS

Suppose let N_t represent the random number of successes in a set interval of "time," $[0, t]$, $t > 0$,

The full collection (family), $N_t, t \geq 0$ is referred to as a counting process

In the contrapositive, let T_n denote the (random) amount of "time" to obtain the fixed # $n \in \mathbb{N}$ of successes.

The corresponding family, $T_n, n \in \mathbb{N}$, is referred to as a waiting-time process

We wait longer than time t to get n successes $(T_n > t) \iff$ the number of successes before time t is less than n $(N_t < n)$

Thus, by definition, we have a fundamental conjugate relation between $(N_t, t \geq 0)$ + $(T_n, n \in \mathbb{N})$

$$T_n > t \iff N_t < n \qquad \forall t \geq 0, n \in \mathbb{N}$$

Suppose $N_t \sim \text{poisson}(\lambda t)$, $t \geq 0$. We can automatically determine the distribution function for each $T_n, n \in \mathbb{N}$

$$F_n(t) = P(T_n \leq t) = P(N_t \geq n) = 1 - P(N_t < n)$$

$$= 1 - e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} \qquad \forall t \geq 0$$

+ immediately a probability density function

$$f_n(t) = F_n'(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \qquad \forall t \geq 0$$

STANDARDIZATION

Let $Z_n = \lambda T_n$, let G_n denote the DF for Z_n. Then $G_n(z) = P(Z_n \leq z) = P(\lambda T_n \leq z) = P(T_n \leq z/\lambda) = F_n(z/\lambda)$

with G_n just as differentiable as F_n, by which we obtain a pdf

$$g_n(z) = G_n'(z) = \frac{1}{\lambda} f_n(z/\lambda) = \frac{1}{(n-1)!} z^{n-1} e^{-z} \qquad \forall z \geq 0$$

But wait, Z_n is just the standard gamma distribution on n degrees of freedom, the waiting time $T_n = \lambda^{-1} Z_n$ w/ rate λ being a scaled version of this standard.

TLDR $\quad N_t \sim \text{poisson}(\lambda t)\ \forall t \geq 0 \implies T_n \sim \text{gamma}(n, \lambda^{-1}) \quad \forall n \in \mathbb{N}$ (converse is also true!)

λ is understood as "average number of successes per unit time"

λ^{-1} is understood as "average amount of time per success"

Proposition 7.0.1 (poisson/gamma reciprocity)

For reciprocal process $\underset{\text{waiting-time}}{T_n > t} \iff \underset{\text{counting}}{N_t < n}$ $\quad \forall t \geq 0, n \in \mathbb{N}$, we have the specific connection in distribution

$$N_t \sim \text{poisson}(\lambda t)\ \forall t \geq 0 \iff T_n \sim \text{gamma}(n, \lambda^{-1}) \quad \forall n \in \mathbb{N}$$

Suppose accidents occur on St George st. at a rate of 0.7 per annum. What is the probability that there will be no accidents in the next two months?

Let the number in time t be denoted N_t so that we may that $N_t \sim \text{poisson}(0.7t)$, then with $t = 1/6$

$$P(N_{1/6} = 0) = e^{-7/60} \frac{(7/60)^0}{0!} = e^{-7/60} \approx 89\%$$

What is the probability that the first accident occurs between 1 + 2 years from now?

$\lambda = 0.7$, $T_1 \sim \text{gamma}(1, 0.7^{-1}) = \exp(0.7^{-1}) \longrightarrow P(1 \le T_1 \le 2) = e^{-0.7} - e^{-1.4} \approx 0.25$

What is the probability that the third accident will occur in the second year, $P(1 \le T_3 \le 2)$?

$T_3 = 0.7^{-1} Z_3$, $Z_3 \sim G(3, 0.7^{-1})$

$$P(1 \le T_3 \le 2) = P(0.7 \le Z_3 \le 1.4) = \int_{0.7}^{1.4} \frac{1}{\Gamma(3)} z^2 e^{-z} \, dz$$

integrate by parts? Nope

$$\begin{aligned} P(1 \le T_3 \le T_2) &= P(T_3 > 1) - P(T_3 > 2) \\ &= P(N_1 < 3) - P(N_2 < 3) \\ &= e^{-0.7}\left(1 + \frac{0.7}{1} + \frac{0.7^2}{2}\right) - e^{-1.4}\left(1 + \frac{1.4}{1} + \frac{1.4^2}{2}\right) \end{aligned}$$

In the general case, we have

$$\begin{aligned} P(s \le T_n \le t) = \int_{\lambda s}^{\lambda t} \frac{1}{\Gamma(n)} z^{n-1} e^{-z} \, dz &= P(T_n > s) - P(T_n > t) \\ &= P(N_s < n) - P(N_t < n) \quad \text{by prop. 7.0.1} \\ &= e^{-\lambda s}\left(1 + \frac{\lambda s}{1} + \cdots + \frac{(\lambda s)^{n-1}}{(n-1)!}\right) - e^{-\lambda t}\left(1 + \frac{\lambda t}{1} + \cdots + \frac{(\lambda t)^{n-1}}{(n-1)!}\right) \end{aligned}$$

The sum of two independent waiting times can be a total waiting time.

The sum of two independent gamma distributions on arbitrary degrees of freedom, $p > 0$ + $q > 0$, remains a gamma distribution, now with $p + q$ df.

Proposition 7.0.2 (sums of independent gammas)

$$Z \sim \text{gamma}(p),\ p > 0,\ W \sim \text{gamma}(q),\ q > 0,\ Z \perp\!\!\!\perp W \implies T = Z + W \sim \text{gamma}(p+q)$$

Corollary 7.0.2.1

$$Z_i \overset{\text{indep}}{\sim} \text{gamma}(p_i),\ i = 1, \ldots, n \implies \sum_{i=1}^{n} Z_i \sim \text{gamma}\left(\sum_{i=1}^{n} p_i\right)$$

In the contex of the poisson process, the waiting time, T_n, for n occurrences is now understood as the independent sum of n waiting times for n successive individual occurrences

$$T_n \overset{d}{=} W_1 + \cdots + W_n \quad \text{w.} \quad W_1, \ldots, W_n \text{ iid } W \sim \exp(\lambda^{-1})$$

With $T = Z + W$ + $U = Z/T$, the joint probability density function for the pair (U, T) automatically separates as a product of individual functions for each component. Unlike Z + T, variables U + T are statistically independent.

Definition 7.0.1 The random variable U is said to have beta distribution on p over q degrees of freedom — denoted $U \sim \text{beta}(p; q)$ or $B(p; q)$ iff

$$U \overset{d}{=} Z/(Z+W) \quad \text{w.} \quad Z \sim G(p),\ W \sim G(q),\ Z \perp\!\!\!\perp W$$

Proposition 7.0.3 (beta-gamma connection)

Given $T = Z + W$, $U = Z/T$, then

$$Z \sim G(p),\ W \sim G(q),\ Z \perp\!\!\!\perp W \iff \begin{cases} (1)\ T \sim G(p+q) \\ (2)\ f_U(u) = \dfrac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} u^{p-1}(1-u)^{q-1},\ 0 < u < 1 \\ (3)\ T \perp\!\!\!\perp U \end{cases}$$

The beta variable U has the interpretation as proportion of time waited. Since $Z = UT$, $U \perp T$, then $\mathbb{E}Z = \mathbb{E}U\,\mathbb{E}T$.

$$\mathbb{E}U = \frac{\mathbb{E}Z}{\mathbb{E}T} = \frac{p}{p+q}$$

$$\mathbb{E}U^2 = \frac{\mathbb{E}Z^2}{\mathbb{E}T^2} = \frac{(p+1)p}{(p+q+1)(p+q)} \qquad \mathbb{E}U^s = \frac{\mathbb{E}Z^s}{\mathbb{E}T^s} = \frac{\Gamma(p+s)/\Gamma(p)}{\Gamma(p+q+s)/\Gamma(p+q)} \quad \text{for } s > -p$$

$$\hookrightarrow \mathbb{E}U^{-1} = \frac{p+q-1}{p-1}$$

OR $U^{-1} = 1 + WZ^{-1} \Rightarrow \mathbb{E}U^{-1} = 1 + \mathbb{E}W\,\mathbb{E}Z^{-1} = 1 + \frac{q}{p-1}$

8 Poisson Process

Theorem 8.0.1 (poisson/gamma reciprocity in poisson process)

For reciprocal processes $T_n > t \iff N_t < n$, $\forall t \geq 0, n \in W$, we have the specific connection

$$N_{t_i} - N_{t_{i-1}} \overset{\text{indep}}{\sim} \text{poisson}(\lambda(t_i - t_{i-1}))\ \forall t_i \uparrow\uparrow \iff T_{n_i} - T_{n_{i-1}} \overset{\text{indep}}{\sim} \text{gamma}(n_i - n_{i-1}, \lambda^{-1})\ \forall 0 < n_i \uparrow\uparrow$$

9 Binomial-Beta Reciprocity

Take an infinite sequence U_i iid $U \sim \text{unif}[0,1]$ of mutually independent standard uniforms, for any specified $n \in \mathbb{N}$, let N_u count the number of U_i's in the sample of the first n, that are $\leq$ specified value $0 < u < 1$.

Then N_u is simply the sum of n iid indicator functions. Wow its binomial

$$N_u = I(U_1 \leq u) + \cdots + I(U_n \leq u) \sim \text{bin}(n, u)$$

Let $U_{(1)}, \ldots, U_{(n)}$ represent $U_1, \ldots, U_n$ in ascending order (i.e. $U_{(1)} = \min\{U_1, \ldots, U_n\}$), then $\forall\, 0 < u < 1$, $k = 1, \ldots, n$, $n \in \mathbb{N}$, we have the simple logical connection

$$U_{(k)} > u \iff N_u < k \quad \equiv \quad U_{(k)} \leq u \iff N_u \geq k$$

Now we can express both the distribution function + probability density function of the k^{th} order statistic, $U_{(k)}$

distribution $\quad F_{(k)}(u) = P(U_{(k)} \leq u) = \sum_{i=1}^{n} \binom{n}{i} u^i (1-u)^{n-i}$

probability density $\quad f_{(k)}(u) = F'(u) = k\binom{n}{k} u^{k-1}(1-u)^{n-k}$

Look familiar? This is because $U_{(k)} \sim \text{beta}(k, n-k+1)$, $k = 1, \ldots, n$

Reconsider any partial binomial sum as an incomplete beta integral

$$\sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{(k-n)!(n-k)!} \int_0^p u^{k-1}(1-u)^{n-k}\, du$$

complement $\quad \sum_{i=0}^{k-1} \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{(k-1)!(n-k)!} \int_p^1 u^{k-1}(1-u)^{n-k}\, du$

10 Normal Distribution

Definition 10.0.1 The random variable Z is said to have a standard normal distribution — denoted $Z \sim N(0,1)$, iff

$$Z \text{ has pdf } \quad \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad \forall z \in \mathbb{R}$$

Definition 10.0.2 The random variable X is said to have a normal distribution w. location $-\infty < \mu < \infty$ & scale $\sigma > 0$

— denoted $X \sim N(\mu, \sigma^2)$ iff $\quad X \overset{d}{=} \mu + \sigma Z \quad$ w. $Z \sim N(0,1)$

$$X \sim N(\mu, \sigma^2) \iff f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

Proposition 10.0.1 (Maxwell's theorem)

(V_1, V_2, V_3) rotationally invariant & statistically independent w. continuous pdf

$$\iff V_1, V_2, V_3 \text{ iid } \sim N(0, \sigma^2) \text{ for some } \sigma > 0$$

Proposition 10.0.2 (normal-gamma connection)

$$Z \sim N(0,1) \iff Z \overset{d}{=} -Z \text{ and } Z^2/2 \sim \text{gamma}(\tfrac{1}{2})$$

For any even value $n = 2k$, using $W = Z^2/2$ gives

$$\mathbb{E}Z^n = 2^k \mathbb{E}W^k = 2^k \frac{\Gamma((2k+1)/2)}{\Gamma(1/2)} = (2k-1)(2k-3)\cdots 1$$

$=$ product of all odd numbers below $n = 2k$

↳ $\mathbb{E}Z^6 = 15$

For any odd value $n = 2k+1$, b/c of reflection symmetry, $-Z^n \overset{d}{=} Z^n$, this means $\mathbb{E}Z^n = 0$

For general normal, $X \overset{d}{=} \mu + \sigma Z$, all power moments are polynomial in μ & σ

↳ $\mathbb{E}X = \mu$, $\mathbb{E}X^2 = \mathbb{E}(\mu + \sigma Z)^2 = \mathbb{E}(\mu^2 + 2\mu\sigma Z + \sigma^2 Z^2) = \mu^2 + \mathbb{E}(\sigma^2 Z^2) = \mu^2 + \sigma^2$

The sum of independent normals is still normal.

Proposition 10.0.3 (sums of independent normals)

For $X_i \overset{\text{indep}}{\sim} N(\mu_i, \sigma_i^2)$ & any constants a_i, $i = 1, 2, \ldots$, & $b \implies \sum_{i=1}^n a_i X_i + b = N\left(\sum_{i=1}^n a_i \mu_i + b, \sum_{i=1}^n a_i^2 \sigma^2\right)$