Statistical Bioinformatics: Makerere
Basics: continued
Timo Koski

TK

02.08.2018

Statistical Bioinformatics: Makerere
Basics: continued
Timo Koski

TK

02.08.2018

KTH Matematik

R.V's, Distributions

Events are

$$\{X = x_i\} \quad : X \text{ assumes the value } x_i$$

$$\{Y = y_j\} \quad : Y \text{ assumes the value } y_j.$$

Then the probability of the event $\{X = x_i\}$ is

$$f_X(x_i) \stackrel{def}{=} P(X = x_i).$$

The probability of the event $\{Y = y_j\}$ is

$$f_Y(y_j) \stackrel{def}{=} P(Y = y_j).$$

KTH Matematik

Furthermore

$$f_X \stackrel{def}{=} (f_X(x_1), \cdots, f_X(x_L))$$

designates a discrete probability distribution on $\mathcal{X}$ and

$$f_Y \stackrel{def}{=} (f_X(y_1), \cdots, f_Y(y_J))$$

designates a discrete probability distribution on $\mathcal{Y}$.

KTH VETENSKAP OCH KONST
KTH Matematik

Joint Probability Distributions

A two dimensional *joint (simultaneous) probability distribution* is a probability defined on the alphabet $\mathcal{X} \times \mathcal{Y}$

$$f_{X,Y}(x_i, y_j) \stackrel{def}{=} P(X = x_i, Y = y_j).$$

Hence $0 \leq f_{X,Y}(x_i, y_j)$ and $\sum_{i=1}^{L} \sum_{j=1}^{J} f_{X,Y}(x_i, y_j) = 1$.

KTH Matematik

Marginal Probability Distributions

Marginal distribution for X:

$$f_X(x_i) = \sum_{j=1}^{J} f_{X,Y}(x_i, y_j).$$

Marginal distribution for Y:

$$f_Y(y_j) = \sum_{i=1}^{L} f_{X,Y}(x_i, y_j).$$

These notions can be extended to n random variables.

KTH Matematik

Conditional Probability Distributions

The conditional probability for $X = x_i$ given $Y = y_j$ is

$$f_{X|Y}(x_i \mid y_j) \overset{def}{=} \frac{f_{X,Y}(x_i, y_j)}{f_Y(y_j)}.$$

The conditional probability for $Y = y_j$ given $X = x_i$ is

$$f_{Y|X}(y_j \mid x_i) \overset{def}{=} \frac{f_{X,Y}(x_i, y_j)}{f_X(x_i)}.$$

Here we assume $f_Y(y_j) > 0$ and $f_X(x_i) > 0$.

KTH Matematik

Conditional Probability Distributions

If for example $f_X(x_i) = 0$, we can make the definition of $f_{Y|X}(y_j \mid x_i)$ arbitrarily through $f_X(x_i) \cdot f_{Y|X}(y_j \mid x_i) = f_{X,Y}(x_i, y_j)$.
In other words

$$f_{Y|X}(y_j \mid x_i) = \frac{\text{prob. for the event } \{X = x_i, Y = y_j\}}{\text{prob. for the event } \{X = x_i\}}.$$

Note that the event $\{X = x_i, Y = y_j\}$ is $\{X = x_i\} \cap \{Y = y_j\}$.

Conditional Probability Distributions

Hence

$$\sum_{i=1}^{L} f_{X|Y}(x_i \mid y_j) = 1,$$

since

$$\sum_{i=1}^{L} f_{X|Y}(x_i \mid y_j) = \frac{\sum_{i=1}^{L} f_{X,Y}(x_i, y_j)}{f_Y(y_j)} = \frac{f_Y(y_j)}{f_Y(y_j)}.$$

Law of Total Probability

$$f_Y(y_j) = \sum_{i=1}^{L} f_{X,Y}(x_i, y_j) = \sum_{i=1}^{L} f_{Y|X}(y_j|x_i) f_X(x_i)$$

Also

$$P(A) = \sum_{l=1}^{k} P\left(A|B_l\right) \cdot P\left(B_l\right),$$

if $A = \cup_{l=1}^{k} A \cap B_l$ and $B_j \cap B_k = \emptyset$, $j \neq k$.

KTH Matematik

Conditional Expectation

Let $\mathcal{X} \overset{g}{\mapsto} R$. Then the conditional expectation of $g(X)$ given $Y = y_j$ is

$$E\left[g(X) \mid Y = y_j\right] = \sum_{i=1}^{L} g(x_i) f_{X|Y}(x_i|y_j)$$

Then we regard $E\left[g(X) \mid Y\right]$ as a r.v. which assumes the value $E\left[g(X) \mid Y = y_j\right]$ when the event $Y = y_j$ occurs.

Conditional Expectation

If $\mathcal{X} \subseteq R$, then $E[X \mid Y]$ is a r.v. with values

$$E[X \mid Y = y_j] = \sum_{i=1}^{L} x_i f_{X|Y}(x_i|y_j).$$

Double Expectation

Thm:

$$E\left[g(X)\right] = E\left[E\left[g(X) \mid Y\right]\right]$$

Proof: $E\left[g(X)\right] = \sum_{i=1}^{L} g(x_i) f_X(x_i)$

$$= \sum_{i=1}^{L} g(x_i) \sum_{j=1}^{J} f_{X,Y}(x_i, y_j) = \sum_{i=1}^{L} g(x_i) \sum_{j=1}^{J} f_{X|Y}(x_i|y_j) f_Y(y_j)$$

$$= \sum_{j=1}^{J} f_Y(y_j) \sum_{i=1}^{L} g(x_i) f_{X|Y}(x_i|y_j) = \sum_{j=1}^{J} f_Y(y_j) E\left[g(X) \mid Y = y_j\right] =$$

$= E\left[E\left[g(X) \mid Y\right]\right]$, as was to be proved.

KTH Matematik

Chain Rule

Let Z be a (discrete) random variable that assumes values in $\mathcal{Z} = \{z_1, z_2, \cdots, z_K\}$. If $f_Z(z_k) > 0$,

$$f_{X,Y|Z}(x_i, y_j \mid z_k) = \frac{f_{X,Y,Z}(x_i, y_j, z_k)}{f_Z(z_k)}.$$

Then we get as an identity

$$f_{X,Y|Z}(x_i, y_j \mid z_k) = \frac{f_{X,Y,Z}(x_i, y_j, z_k)}{f_{Y,Z}(y_j, z_k)} \cdot \frac{f_{Y,Z}(y_j, z_k)}{f_Z(z_k)}$$

and by definition of conditional probability

$$= f_{X|Y,Z}(x_i \mid y_j, z_k) \cdot f_{Y|Z}(y_j \mid z_k).$$

KTH Matematik

Chain Rule

In other words,

$$f_{X,Y|Z}(x_i, y_j \mid z_k) = f_{X|Y,Z}(x_i \mid y_j, z_k) \cdot f_{Y|Z}(y_j \mid z_k).$$

This is the chain rule.

Independence

X and Y are *independent* random variables if and only if

$$f_{X,Y}(x_i, y_j) = f_X(x_i) \cdot f_Y(y_j)$$

for *all pairs* (x_i, y_j) in $\mathcal{X} \times \mathcal{Y}$. In other words all events $\{X = x_i\}$ and $\{Y = y_j\}$ are to be independent.

KTH Matematik

Independence

X and Y are *independent* random variables if and only if

$$f_{X,Y}(x_i, y_j) = f_X(x_i) \cdot f_Y(y_j)$$

for *all pairs* (x_i, y_j) in $\mathcal{X} \times \mathcal{Y}$. In other words all events $\{X = x_i\}$ and $\{Y = y_j\}$ are to be independent.

Independence of Functions of Indpendent R.V's

If $X_1, X_2, \ldots, X_n$ are independent random variables, then Y_1, $Y_2, \ldots$, Y_n defined by

$$Y_i = g_i\left(X_i\right), i = 1, \ldots, n$$

respectively, are independent random variables for any functions $g_1(x), \ldots g_n(x)$.

KTH Matematik

$E[X \cdot Y]$

X and Y are independent and with a finite of real numbers as alphabet

$$E[X \cdot Y] = \sum_{i=1}^{L} \sum_{j=1}^{K} x_i \cdot y_j f_{X,Y}(x_i, y_j) = \sum_{i=1}^{L} \sum_{j=1}^{K} x_i \cdot y_j f_X(x_i) f_Y(y_j)$$

$$= \sum_{i=1}^{L} x_i f_X(x_i) \sum_{j=1}^{K} y_j f_Y(y_j) = E[X] \cdot E[Y]$$

KTH Matematik

Sequence

A *sequence* or a *string* $\mathbf{x}$ is an ordered list of m symbols from an alphabet $\mathcal{X}$ written contiguously from left to right

$$\mathbf{x} = x_{l_1} x_{l_2} \ldots x_{l_i} \ldots x_{l_m}; x_{l_i} \in \mathcal{X}, i = 1, \ldots, m.$$

KTH Matematik

Probabilistic Models

X_i are (**discrete**) random variables that assume values in $\mathcal{X}$

$$\mathbf{x} = x_{l_1} x_{l_2} \ldots x_{l_i} \ldots x_{l_m}; x_{l_i} \in \mathcal{X}, i = 1, \ldots, m.$$

Assign a probability (a joint distribution)

$$P(\mathbf{x}) = P(X_1 = x_{l_1}, \ldots, X_m = x_{l_m})$$

to a string. We need a model.

KTH Matematik

Modelling by Independence

$$\mathbf{x} = x_{l_1} x_{l_2} \ldots x_{l_i} \ldots x_{l_m} ; x_{l_i} \in \mathcal{X}, i = 1, \ldots, m.$$

We assign a probability (a joint distribution) by

$$P(\mathbf{x}) = P\left(X_1 = x_{l_1}, \cdots, X_m = x_{l_m}\right)$$

$$= \prod_{i=1}^{m} P\left(X_i = x_{l_i}\right) = \prod_{i=1}^{m} f_{X_i}\left(x_{l_i}\right).$$

KTH Matematik

Example: Bernoulli R.V's

X_i is a random variable assuming values in $\{0, 1\}$

$$f_X(1) = P(X = 1) = p, q = P(X = 0) = 1 - p.$$

$$X_i \in Be(p), i = 1, \ldots, m.$$

I.I.D (=independent, identically distributed) Bernoulli R.V.s.

$$P(X_1 = x_{l_1}, \ldots, X_m = x_{l_m}) = p^k \cdot (1 - p)^{m-k}$$

where $k = \sum_{i=1}^{m} X_i$ (=the number of ones in $x_{l_1} x_{l_2} \ldots x_{l_i} \ldots x_{l_m}$).

KTH Matematik

Modelling by Chain Rule

$$P(\mathbf{x}) = P\left(X_1 = x_{l_1}, \cdots, X_m = x_{l_m}\right) =$$

$$= \prod_{i=1}^{m} P\left(X_i = x_{l_i} \mid X_1 = x_{l_1} \ldots X_{i-1} = x_{l_{i-1}}\right)$$

where

$$P\left(X_1 = x_{l_1} \mid X_0 = x_{l_0}\right) = P\left(X_1 = x_{l_1}\right).$$

This is for obvious reasons unpractical.

Conditional Independence

The random variables X, Y are called *conditionally independent* given Z if

$$f_{X,Y|Z}(x_i, y_j|z_k) = f_{X|Z}(x_i|z_k) \cdot f_{Y|Z}(y_j|z_k)\cdot$$

for all triples $(z_k, x_i, y_j) \in \mathcal{Z} \times \mathcal{X} \times \mathcal{Y}$.
Later: the Markov property.

KTH Matematik

Multinomial Probability Distribution

Let $X_1, X_2, \ldots, X_n$ be I.I.D. random variables assuming values in

$$\mathcal{X} = \{x_1, \cdots, x_L\}$$

with the common distribution

$$p_l = P\left(X_i = x_l\right), l = 1, 2, \ldots, L.$$

Then

$$P\left(X_1 = x_{i_1}, X_2 = x_{i_2}, \ldots, X_n = x_{i_n}\right) = p_{i_1} \cdot p_{i_2} \cdots p_{i_n}.$$

KTH Matematik

Multinomial Probability Distribution

Let for $l = 1, 2, \ldots, L$

$$n_l = \text{ the number of times the symbol } x_l \text{ is}$$

$$\text{found in } \mathbf{x} = x_{i_1} x_{i_2} \ldots x_{i_n}.$$

Thus $n_1 + n_2 + \ldots + n_L = n$. Then

$$P(\mathbf{x}) = P(X_1 = x_{i_1}, X_2 = x_{i_2}, \ldots, X_n = x_{i_n}) = p_1^{n_1} \cdot p_2^{n_2} \cdots p_L^{n_L}.$$

KTH Matematik

Multinomial Probability Distribution

If we introduce the random variables $Y_1, Y_2, \ldots, Y_L$ as functions of the random variables $X_1, \ldots, X_n$ by

$$Y_l = \text{ the number of times } X_i = x_l,\ i = 1, 2, \ldots, n,$$

then the joint distribution of $Y_1, Y_2, \ldots, Y_L$ is

$$P\left(Y_1 = n_1, Y_2 = n_2, \ldots, Y_L = n_L\right) = C \cdot p_1^{n_1} \cdots p_L^{n_L}.$$

KTH Matematik

Multinomial Probability Distribution

$$P\left(Y_1 = n_1, Y_2 = n_2, \ldots, Y_L = n_L\right) = C \cdot p_1^{n_1} \cdots p_L^{n_L}.$$

is by independence (symmetry) just the sum of probabilities of all those outcomes of $X_1, \ldots, X_n$ that have exactly $n_1, n_2, \ldots, n_L$ as their frequency counts. Therefore a combinatorial argument shows that

$$C = \frac{n!}{n_1! n_2! \ldots n_L!}, \qquad (1)$$

which is called the *multinomial coefficient*.

KTH Matematik

Multinomial Probability Distribution

The probability

$$P\left(Y_1 = n_1, Y_2 = n_2, \ldots, Y_L = n_L\right) = \frac{n!}{n_1! n_2! \ldots n_L!} \cdot p_1^{n_1} \cdots p_L^{n_L}.$$

is called the *multinomial distribution*. The binomial distribution ($L = 2$) is a special case. There is a *Whittle distribution* generalizing the multinomial distribution for Markov chains.
$Y_1, Y_2, \ldots, Y_L$ are dependent random variables.

Multinomial Probability Distribution

Each Y_i considered on its own has a binomial distribution $Y_i \in \mathrm{Bin}\left(n, p_i\right)$.

KTH Matematik

We can regard the probability assignment

$$P(\mathbf{x}) = p_1^{n_1} \cdot p_2^{n_2} \cdots p_L^{n_L},$$

as a statement of conditional independence

$$P(\mathbf{x} \mid \underline{p}) =$$

$$P(X_1 = x_{i_1}, X_2 = x_{i_2}, \ldots, X_n = x_{i_n} \mid \underline{p}) = p_1^{n_1} \cdot p_2^{n_2} \cdots p_L^{n_L},$$

where $\underline{p} = (p_1, p_2, \cdots, p_L)$ is seen as an outcome of a continuous random variable.

Probability Generating Function (PGF)

$$G(t) = E\left(t^X\right) = \sum_{k=0}^{\infty} t^k f_X(k).$$

a.k.a (at least in this course) the Mellin transform.

KTH VETENSKAP OCH KONST
KTH Matematik

- $$\frac{d}{dt}G(1) = \sum_{k=1}^{\infty} kt^{k-1} f_X(k) \mid_{t=1}$$
  $$= E[X]$$
- $S = X + Y$, X and Y integer valued, independent,
  $$G_S(t) = E\left(t^S\right) =$$
  $$E\left(t^{X+Y}\right) = E\left(t^X\right) \cdot E\left(t^Y\right) = G_X(t) \cdot G_Y(t),$$
  since functions of independent random variables are independent.

KTH Matematik

PGF : Properties

$Z = X_1 + X_2 + \ldots + X_n$, X_i integer valued, independent,

$$G_Z(t) = E\left(t^Z\right) =$$

$$E\left(t^{X_1+X_2+\ldots+X_n}\right) = E\left(t^{X_1}t^{X_2}\cdots t^{X_n}\right) =$$

$$E\left(t^{X_1}\right)\cdot E\left(t^{X_2}\right)\cdots E\left(t^{X_n}\right) = G_{X_1}(t)\cdot G_{X_2}(t)\cdots G_{X_n}(t)$$

X_i integer valued, I.I.D.,

$$G_Z(t) = \left(G_X(t)\right)^n.$$

PGF : Examples

- $X \in \mathrm{Be}(p)$

$$G_X(t) = 1 - p + pt.$$

- $Y \in \mathrm{Bin}(n, p)$

$$G_Y(t) = (1 - p + pt)^n$$

Moment Generating Function

$$\phi_X(s) \stackrel{def}{=} E\left(e^{sX}\right) = \begin{cases} \sum_{x_i} e^{sx_i} f_X(x_i) & X \text{ discrete} \\ \int_{-\infty}^{\infty} e^{sx} f_X(x) dx & X \text{ continuous} \end{cases}$$

is the moment generating function (MGF).

MGF: Properties

- $$\frac{d}{ds}\phi(0) = E\left[X\right]$$
- $$\phi(0) = 1$$
- $$\frac{d^k}{ds^k}\phi(0) = E\left[X^k\right].$$

KTH VETENSKAP OCH KONST
KTH Matematik

MGF: Properties

$S_n = X_1 + X_2 + \ldots + X_n$, X_i independent.

$$\phi_{S_n}(s) = E\left(e^{sS_n}\right) =$$

$$E\left(e^{s(X_1+X_2+\ldots+X_n)}\right) = E\left(e^{sX_1}e^{sX_2}\cdots e^{sX_n}\right) =$$

$$E\left(e^{sX_1}\right)E\left(e^{sX_2}\right)\cdots E\left(e^{sX_n}\right) = \phi_{X_1}(s)\cdot\phi_{X_2}(s)\cdots\phi_{X_n}(s)$$

X_i I.I.D.,

$$\phi_{S_n}(s) = \left(\phi_X(s)\right)^n.$$

KTH Matematik

MGF: Properties

$S_n = X_1 + X_2 + \ldots + X_n$, X_i independent, $E(X_i) = \mu_i$.

$$\phi_{S_n}(s) = \phi_{X_1}(s) \cdot \phi_{X_2}(s) \cdots \phi_{X_n}(s)$$

$$E[S_n] = \frac{d}{ds}\phi_{S_n}(0) = \sum_{i=1}^{n} \phi_{X_1}(0) \cdots \frac{d}{ds}\phi_{X_i}(0) \cdots \phi_{X_n}(0) =$$

$$= \sum_{i=1}^{n} \mu_i.$$

KTH Matematik

Indicators

A is an event and X is a random variable

$$I_A(X) = \begin{cases} 1 & \text{if } X \text{ hits } A \\ 0 & \text{otherwise.} \end{cases}$$

X_1, X_2 $\ldots X_n$ are independent R.V's, we consider

$$\sum_{i=1}^{n} I_A(X_i).$$

KTH Matematik

Indicators

From the preceding

$$E\left[\sum_{i=1}^{n} I_A(X_i)\right] = \sum_{i=1}^{n} E\left[I_A(X_i)\right]$$

$$= \sum_{i=1}^{n} P_{X_i}(A).$$

MGF of a Random Walk

$X_i = +1$ with probability p, $X_i = -1$ with probability $q = 1 - p$, indpendent.

$$S_n = X_1 + X_2 + \ldots + X_n, S_0 = 0$$

is called a 'random walk'. The moment generating function of one random step $S = S_1 = X_1$ is

$$\phi_S(s) = pe^s + qe^{-s}$$

Then

$$\phi_{S_n}(s) = \left(pe^s + qe^{-s}\right)^n$$

KTH Matematik

MGF of a Random Walk

$S_2 = -2$ with prob. q^2, $S_2 = 0$ with prob. $2pq$ and $S_2 = 2$ with prob. p^2. By definition of m.g.f. we have

$$q^2 e^{-2s} + 2pq + p^2 e^{2s}.$$

$S_2 = -2$ with prob. q^2, $S_2 = 0$ with prob. $2pq$ and $S_2 = 2$ with prob. p^2. By definition of m.g.f. we have

$$q^2 e^{-2s} + 2pq + p^2 e^{2s},$$

which equals

$$\left(pe^s + qe^{-s}\right)^2 = \phi_{S_2}(s)$$

from the above.

KTH Matematik

A Thm for Bioinformatics

Let X be a discrete R.V. with moment generating function $\phi_X(s)$. Say that X can take at least one negative value (say $-a$) with positive probability $f_X(-a)$ and at least one positive value (say b) with positive probability $f_X(b)$, and that the expectation of X is nonzero. Then there exists a unique nonzero value s^* such that

$$\phi_X(s^*) = 1.$$

KTH Matematik

Proof (1)

Proof: Assume $\phi_X(s)$ is defined for all s.

$$\phi_X(s) > f_X(-a)e^{-as}, \phi_X(s) > f_X(b)e^{bs}$$

Hence $\phi_X(s) \to \infty$, as $s \to \infty$ and also as $s \to -\infty$.

KTH Matematik

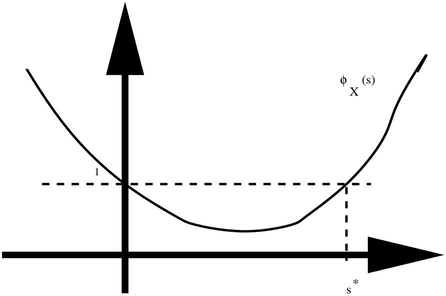
Proof (2)

$$\frac{d^2}{ds^2}\phi(s) = \sum x^2 e^{sx} f_X(x) \geq 0.$$

Hence $\phi(s)$ is convex as function of s. $\phi(0) = 1$ and the the expectation $\frac{d}{ds}\phi(0) = E[X]$ is nonzero by assumption. If $E[X]$ is negative, then the graph of $\phi_X(s)$ must be as in the Figure below:

KTH Matematik

Proof (3)



Similarly for $E[X]$ positive. □