LibreTexts

5.22: Discrete Uniform Distributions

Uniform Distributions on a Finite Set

Suppose that S is a nonempty, finite set. A random variable X taking values in S has the *uniform distribution* on S if

$$\mathbb{P}(X \in A) = \frac{\#(A)}{\#(S)}, \quad A \subseteq S \tag{5.22.1}$$

The discrete uniform distribution is a special case of the general uniform distribution with respect to a measure, in this case counting measure. The distribution corresponds to picking an element of S *at random*. Most classical, combinatorial probability models are based on underlying discrete uniform distributions. The chapter on Finite Sampling Models explores a number of such models.

The probability density function f of X is given by

$$f(x) = \frac{1}{\#(S)}, \quad x \in S \tag{5.22.2}$$

Proof

This follows from the definition of the (discrete) probability density function: $\mathbb{P}(X \in A) = \sum_{x \in A} f(x)$ for $A \subseteq S$. Or more simply, $f(x) = \mathbb{P}(X = x) = 1/\#(S)$.

Like all uniform distributions, the discrete uniform distribution on a finite set is characterized by the property of constant density on the set. Another property that all uniform distributions share is invariance under conditioning on a subset.

Suppose that R is a nonempty subset of S. Then the conditional distribution of X given $X \in R$ is uniform on R.

Proof

For $A \subseteq R$,

$$\mathbb{P}(X \in A \mid X \in R) = \frac{\mathbb{P}(X \in A)}{\mathbb{P}(X \in R)} = \frac{\#(A)/\#(S)}{\#(R)/\#(S)} = \frac{\#(A)}{\#(R)} \tag{5.22.3}$$

If $h: S \to \mathbb{R}$ then the expected value of $h(X)$ is simply the arithmetic average of the values of h:

$$\mathbb{E}[h(X)] = \frac{1}{\#(S)} \sum_{x \in S} h(x) \tag{5.22.4}$$

Proof

This follows from the change of variables theorem for expected value:

$$\mathbb{E}[h(X)] = \sum_{x \in S} f(x) h(x) = \frac{1}{\#(S)} \sum_{x \in S} h(x) \tag{5.22.5}$$

The entropy of X depends only on the number of points in S.

The entropy of X is $H(X) = \ln[\#(S)]$.

Proof

Let $n = \#(S)$. Then

$$H(X) = \mathbb{E}\{-\ln[f(X)]\} = \sum_{x \in S} -\ln\left(\frac{1}{n}\right)\frac{1}{n} = -\ln\left(\frac{1}{n}\right) = \ln(n) \tag{5.22.6}$$

Uniform Distributions on Finite Subsets of $\mathbb{R}$

Without some additional structure, not much more can be said about discrete uniform distributions. Thus, suppose that $n \in \mathbb{N}_+$ and that $S = \{x_1, x_2, \ldots, x_n\}$ is a subset of $\mathbb{R}$ with n points. We will assume that the points are indexed in order, so that $x_1 < x_2 < \cdots < x_n$. Suppose that X has the uniform distribution on S.

The probability density function f of X is given by $f(x) = \frac{1}{n}$ for $x \in S$.

The distribution function F of X is given by

1. $F(x) = 0$ for $x < x_1$
2. $F(x) = \frac{k}{n}$ for $x_k \le x < x_{k+1}$ and $k \in \{1, 2, \ldots n-1\}$
3. $F(x) = 1$ for $x > x_n$

Proof

This follows from the definition of the distribution function: $F(x) = \mathbb{P}(X \le x)$ for $x \in \mathbb{R}$.

The quantile function F^{-1} of X is given by $F^{-1}(p) = x_{\lceil np \rceil}$ for $p \in (0, 1]$.

Proof

By definition, $F^{-1}(p) = x_k$ for $\frac{k-1}{n} < p \le \frac{k}{n}$ and $k \in \{1, 2, \ldots, n\}$. It follows that $k = \lceil np \rceil$ in this formulation.

The moments of X are ordinary arithmetic averages.

For $k \in \mathbb{N}$

$$\mathbb{E}\left(X^k\right) = \frac{1}{n} \sum_{i=1}^n x_i^k \tag{5.22.7}$$

In particular,

The mean and variance of X are

1. $\mu = \frac{1}{n} \sum_{i=1}^n x_i$
2. $\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$

Uniform Distributions on Discrete Intervals

We specialize further to the case where the finite subset of $\mathbb{R}$ is a *discrete interval*, that is, the points are uniformly spaced.

The Standard Distribution

Suppose that $n \in \mathbb{N}_+$ and that Z has the discrete uniform distribution on $S = \{0, 1, \ldots, n-1\}$. The distribution of Z is the *standard discrete uniform distribution* with n points.

Of course, the results in the previous subsection apply with $x_i = i - 1$ and $i \in \{1, 2, \ldots, n\}$.

The probability density function g of Z is given by $g(z) = \frac{1}{n}$ for $z \in S$.

Open the Special Distribution Simulation and select the discrete uniform distribution. Vary the number of points, but keep the default values for the other parameters. Note the graph of the probability density function. Run the simulation 1000 times and compare the empirical density function to the probability density function.

The distribution function G of Z is given by $G(z) = \frac{1}{n}(\lfloor z \rfloor + 1)$ for $z \in [0, n-1]$.

Proof

Note that $G(z) = \frac{k}{n}$ for $k - 1 \le z < k$ and $k \in \{1, 2, \ldots n-1\}$. Thus $k - 1 = \lfloor z \rfloor$ in this formulation.

The quantile function G^{-1} of Z is given by $G^{-1}(p) = \lceil np \rceil - 1$ for $p \in (0, 1]$. In particular

1. $G^{-1}(1/4) = \lceil n/4 \rceil - 1$ is the first quartile.
2. $G^{-1}(1/2) = \lceil n/2 \rceil - 1$ is the median.
3. $G^{-1}(3/4) = \lceil 3n/4 \rceil - 1$ is the third quartile.

Proof

Note that $G^{-1}(p) = k - 1$ for $\frac{k-1}{n} < p \le \frac{k}{n}$ and $k \in \{1, 2, \ldots, n\}$. Thus $k = \lceil np \rceil$ in this formulation.

Open the special distribution calculator and select the discrete uniform distribution. Vary the number of points, but keep the default values for the other parameters. Note the graph of the distribution function. Compute a few values of the distribution function and the quantile function.

For the standard uniform distribution, results for the moments can be given in closed form.

The mean and variance of Z are

1. $\mathbb{E}(Z) = \frac{1}{2}(n-1)$
2. $\operatorname{var}(Z) = \frac{1}{12}(n^2 - 1)$

Proof

Recall that

$$\sum_{k=0}^{n-1} k = \frac{1}{2} n(n-1) \tag{5.22.8}$$

$$\sum_{k=0}^{n-1} k^2 = \frac{1}{6} n(n-1)(2n-1) \tag{5.22.9}$$

Hence $\mathbb{E}(Z) = \frac{1}{2}(n-1)$ and $\mathbb{E}(Z^2) = \frac{1}{6}(n-1)(2n-1)$. Part (b) follows from $\operatorname{var}(Z) = \mathbb{E}(Z^2) - [\mathbb{E}(Z)]^2$.

Open the Special Distribution Simulation and select the discrete uniform distribution. Vary the number of points, but keep the default values for the other parameters. Note the size and location of the mean$\pm$standard devation bar. Run the simulation 1000 times and compare the empirical mean and standard deviation to the true mean and standard deviation.

The skewness and kurtosis of Z are

1. $\operatorname{skew}(Z) = 0$
2. $\operatorname{kurt}(Z) = \frac{3}{5} \frac{3n^2 - 7}{n^2 - 1}$

Proof

Recall that

$$\sum_{k=1}^{n-1} k^3 = \frac{1}{4}(n-1)^2 n^2 \tag{5.22.10}$$

$$\sum_{k=1}^{n-1} k^4 = \frac{1}{30}(n-1)(2n-1)(3n^2 - 3n - 1) \tag{5.22.11}$$

Hence $\mathbb{E}(Z^3) = \frac{1}{4}(n-1)^2 n$ and $\mathbb{E}(Z^4) = \frac{1}{30}(n-1)(2n-1)(3n^2 - 3n - 1)$. The results now follow from the results on the mean and varaince and the standard formulas for skewness and kurtosis. Of course, the fact that $\operatorname{skew}(Z) = 0$ also follows from the symmetry of the distribution.

Note that $\operatorname{skew}(Z) \to \frac{9}{5}$ as $n \to \infty$. The limiting value is the skewness of the uniform distribution on an interval.

Z has probability generating function P given by $P(1) = 1$ and

$$P(t) = \frac{1}{n} \frac{1 - t^n}{1 - t}, \quad t \in \mathbb{R} \setminus \{1\} \tag{5.22.12}$$

Proof

$$P(t) = \mathbb{E}\left(t^Z\right) = \frac{1}{n} \sum_{k=0}^{n-1} t^k = \frac{1}{n} \frac{1 - t^n}{1 - t} \tag{5.22.13}$$

The General Distribution

We now generalize the standard discrete uniform distribution by adding location and scale parameters.

Suppose that Z has the standard discrete uniform distribution on $n \in \mathbb{N}_+$ points, and that $a \in \mathbb{R}$ and $h \in (0, \infty)$. Then $X = a + hZ$ has the *uniform distribution* on n points with *location parameter* a and *scale parameter* h.

Note that X takes values in

$$S = \{a, a+h, a+2h, \ldots, a+(n-1)h\} \tag{5.22.14}$$

so that S has n elements, starting at a, with step size h, a *discrete interval*. In the further special case where $a \in \mathbb{Z}$ and $h = 1$, we have an *integer interval*. Note that the last point is $b = a + (n-1)h$, so we can clearly also parameterize the distribution by the endpoints a and b, and the step size h. With this parametrization, the number of points is $n = 1 + (b-a)/h$. For the remainder of this discussion, we assume that X has the distribution in the definiiton. Our first result is that the distribution of X really is uniform.

X has probability density function f given by $f(x) = \frac{1}{n}$ for $x \in S$

Proof

Recall that $f(x) = g\left(\frac{x-a}{h}\right)$ for $x \in S$, where g is the PDF of Z.

Open the Special Distribution Simulation and select the discrete uniform distribution. Vary the parameters and note the graph of the probability density function. For various values of the parameters, run the simulation 1000 times and compare the empirical density function to the probability density function.

The distribution function F of x is given by

$$F(x) = \frac{1}{n}\left(\left\lfloor \frac{x-a}{h} \right\rfloor + 1\right), \quad x \in [a, b] \tag{5.22.15}$$

Proof

Recall that $F(x) = G\left(\frac{x-a}{h}\right)$ for $x \in S$, where G is the CDF of Z.

The quantile function F^{-1} of X is given by $G^{-1}(p) = a + h\left(\lceil np \rceil - 1\right)$ for $p \in (0, 1]$. In particular

1. $F^{-1}(1/4) = a + h\left(\lceil n/4 \rceil - 1\right)$ is the first quartile.
2. $F^{-1}(1/2) = a + h\left(\lceil n/2 \rceil - 1\right)$ is the median.
3. $F^{-1}(3/4) = a + h\left(\lceil 3n/4 \rceil - 1\right)$ is the third quartile.

Proof

Recall that $F^{-1}(p) = a + hG^{-1}(p)$ for $p \in (0, 1]$, where G^{-1} is the quantile function of Z.

Open the special distribution calculator and select the discrete uniform distribution. Vary the parameters and note the graph of the distribution function. Compute a few values of the distribution function and the quantile function.

The mean and variance of X are

1. $\mathbb{E}(X) = a + \frac{1}{2}(n-1)h = \frac{1}{2}(a+b)$

2. $\text{var}(X) = \frac{1}{12}(n^2 - 1)h^2 = \frac{1}{12}(b-a)(b-a+2h)$

Proof

Recall that $\mathbb{E}(X) = a + h\mathbb{E}(Z)$ and $\text{var}(X) = h^2\text{var}(Z)$, so the results follow from the corresponding results for the standard distribution.

Note that the mean is the average of the endpoints (and so is the midpoint of the interval $[a, b]$) while the variance depends only on the number of points and the step size.

Open the Special Distribution Simulator and select the discrete uniform distribution. Vary the parameters and note the shape and location of the mean/standard deviation bar. For selected values of the parameters, run the simulation 1000 times and compare the empirical mean and standard deviation to the true mean and standard deviation.

The skewness and kurtosis of Z are

1. $\text{skew}(X) = 0$
2. $\text{kurt}(X) = \frac{3}{5}\frac{3n^2-7}{n^2-1}$

Proof

Recall that skewness and kurtosis are defined in terms of the standard score, and hence are the skewness and kurtosis of X are the same as the skewness and kurtosis of Z.

X has moment generating function M given by $M(0) = 1$ and

$$M(t) = \frac{1}{n}e^{ta}\frac{1-e^{nth}}{1-e^{th}}, \quad t \in \mathbb{R} \setminus \{0\} \tag{5.22.16}$$

Proof

Note that $M(t) = \mathbb{E}\left(e^{tX}\right) = e^{ta}\mathbb{E}\left(e^{thZ}\right) = e^{ta}P\left(e^{th}\right)$ where P is the probability generating function of Z.

Related Distributions

Since the discrete uniform distribution on a discrete interval is a location-scale family, it is trivially closed under location-scale transformations.

Suppose that X has the discrete uniform distribution on $n \in \mathbb{N}_+$ points with location parameter $a \in \mathbb{R}$ and scale parameter $h \in (0, \infty)$. If $c \in \mathbb{R}$ and $w \in (0, \infty)$ then $Y = c + wX$ has the discrete uniform distribution on n points with location parameter $c + wa$ and scale parameter wh.

Proof

By definition we can take $X = a + hZ$ where Z has the standard uniform distribution on n points. Then $Y = c + wX = (c + wa) + (wh)Z$.

In terms of the endpoint parameterization, X has left endpoint a, right endpoint $a + (n-1)h$, and step size h while Y has left endpoint $c + wa$, right endpoint $(c + wa) + (n-1)wh$, and step size wh.

The uniform distribution on a discrete interval converges to the continuous uniform distribution on the interval with the same endpoints, as the step size decreases to 0.

Suppose that X_n has the discrete uniform distribution with endpoints a and b, and step size $(b-a)/n$, for each $n \in \mathbb{N}_+$. Then the distribution of X_n converges to the continuous uniform distribution on $[a, b]$ as $n \to \infty$.

Proof

The CDF F_n of X_n is given by

$$F_n(x) = \frac{1}{n}\left\lfloor n\frac{x-a}{b-a}\right\rfloor, \quad x \in [a, b] \tag{5.22.17}$$

But $ny - 1 \le \lfloor ny \rfloor \le ny$ for $y \in \mathbb{R}$ so $\lfloor ny \rfloor / n \to y$ as $n \to \infty$. Hence $F_n(x) \to (x - a)/(b - a)$ as $n \to \infty$ for $x \in [a, b]$, and this is the CDF of the continuous uniform distribution on $[a, b]$.