Ex: Bucket Sorting (Gonnet, 4.2.3, pg. 176) (1-pass version)
Assume you know the distribution of a set of keys to be sorted. That is, assume we have \( M \) buckets that partition the key space, \( M \) keys, and \( \Pr\{\text{key belongs in bucket } b_m\} = 1/M, 1 \leq m \leq M \).

\[
\text{for } m := 1 \text{ to } M \text{ do initialize bucket } b_m \quad \theta(M),
\]

\[
\text{for each key do INSERT(key, appropriate bucket) } \quad \theta(M)
\]

InsertionSort each bucket \( \theta(\text{length of bucket}) \)

Concatenate buckets 1,2,...\( M \) \( \theta(M) \)

InsertionSort requires time quadratic in # items in bucket.

Let \( n_m, 1 \leq m \leq M \), be a r.v. denoting the number of keys in bucket \( b_m \). Total expected time to InsertionSort the \( M \) buckets:

\[
\mathbb{E}\left[ n_m^2 \right] = \sum_{1 \leq m \leq M} \mathbb{E}\left[ n_m^2 \right] = \mathbb{V}(n_m) + (\mathbb{E}[n_m])^2
\]

What is \( \mathbb{P}(n_b = k) \)? 0 for \( k < 0 \) or \( k > M \).

Binomial Distribution: \( \mathbb{P}(n_b = k) = \binom{M}{k} p^k q^{M-k} \)

Note that it’s legal since \( \mathbb{P}(n_b = k) = \binom{M}{k} p^k q^{M-k} \geq 0 \) for all \( k \) and

\[
\sum_k \binom{M}{k} p^k q^{M-k} = \sum_{0 \leq k \leq M} \binom{M}{k} p^k q^{M-k} = (p+q)^M = 1
\]

\[
\mathbb{E}[n_b] = \sum_{0 \leq k \leq M} k \cdot \mathbb{P}(n_b = k) = \sum_{0 \leq k \leq M} k \binom{M}{k} p^k q^{M-k}
\]

Get rid of \( k \) multiplier: Absorption/extraction

\[
\mathbb{E}[n_b] = \sum_{0 \leq k \leq M} M \binom{M-1}{k-1} p^k q^{M-k} = M \sum_{0 \leq k \leq M} \binom{M-1}{k-1} p^k q^{M-k}
\]

(make exponent of \( p \) match lower term of binomial)

\[
= M p \sum_{0 \leq k \leq M} \binom{M-1}{k-1} p^{k-1} q^{M-k}
\]

(replace \( k \) by \( k+1 \))

\[
= M p \sum_{0 \leq k+1 \leq M} \binom{M-1}{k} p^k q^{M-1-k} = M p \sum_{-1 \leq k \leq M} \binom{M-1}{k} p^k q^{M-1-k}
\]
\[
M_0 \sum_{0 \leq k \leq M} \binom{M - 1}{k} p^k q^{M - 1 - k} = M_0 \sum_{0 \leq k \leq M} \binom{M - 1}{k} p^k q^{M - 1 - k} = M_0 (p + q)^{M - 1} = M p
\]

Another approach: \( n_b = \sum_{1 \leq m \leq M} X_m \), so
\[
E[n_b] = E \left[ \sum_{1 \leq m \leq M} X_m \right] = \sum_{1 \leq m \leq M} E[X_m] = \sum_{1 \leq m \leq M} p = M p
\]

\( V(n_m) = V(X_m) = E[X_m^2] - (E[X_m])^2 \)
\[
E[X_m^2] = 1^2 p + 0^2 q = p, \quad (E[X_m])^2 = p^2, \quad V(X_m) = p - p^2 = p(1 - p) = pq
\]

**Theorem**: If the \( M \) r.v.s are independent, then
\[
V \left( \sum_{1 \leq m \leq M} X_m \right) = \sum_{1 \leq m \leq M} V(X_m)
\]

So for Bucket Sort, \( V \left( \sum_{1 \leq m \leq M} X_m \right) = Mpq \), and the expected time to Insertion Sort bucket \( b_m = E[n_m^2] = V(n_m) + (E[n_m])^2 = Mpq + Mp = Mp(2 - p) \)

Noting that \( p = \frac{1}{M} \), this is \( 2 - \frac{1}{M} = \text{constant time to Insertion Sort each bucket, \\& (expected) linear time to Bucket Sort the array.} \)

**Ex**: Assume \( n \) bits to be transmitted across a channel, \( \text{Pr\{error-free transmission of a bit\}} = p \), \& errors are independent. If code capable of detecting/correcting up to \( e \) errors, then probability of correct transmission of the \( n \) bits =
\[
\sum_{0 \leq k \leq e} \binom{n}{k} q^k p^{n-k}
\]

**Ex**: Choosing a leader in a ring  Itai \& Rodeh, *Symmetry Breaking in Distributive Networks*, 22\textsuperscript{nd} FOCS, 1981

Given a ring of \( n \) identical processors, in the same state, but each processor endowed with a (non-psuedo) random number generator, they must all agree on a leader. Each processor can only transmit 1-bit message each time. Initially, all processors are *active*.

**loop**  Each processor draws from \( U(1..\#\text{active processors}) \) independently of the others. All processors that chose 1 send this message.

if 0 processors chose 1 then repeat the loop
if 1 processor chose 1 then it's the leader
if >1 processors chose 1 then processors that didn't choose 1 become *inactive* & the loop is repeated

Let \( \text{Pr}(k \text{ processors chose } 1) = \pi(n,k) = \binom{n}{k} \left( \frac{1}{n} \right)^k \left( \frac{n-1}{n} \right)^{n-k} \)

Let \( \beta(n) - \text{expected number of passes through the loop} \)
\[ \beta(n) = 1 + \pi(n, 0) \beta(n) + \sum_{2 \leq k \leq n} \pi(n, k) \beta(k) \Rightarrow \beta(n) = \frac{1 + \sum \pi(n, k) \beta(k)}{1 - \pi(n, 0) - \pi(n, n)} \]

...Results: For all \( n \geq 2 \), \( \beta(n) < e = 2.718 \ldots \)

**Ex:** Counting distinct words Flajolet & Martin, "Probabilistic Counting Algorithms for Database Applications", *J. Comp. Syst. Sciences*, 1985

Assume there are \( n \) (not necessarily distinct) words in a text. Let there be a hashing function \( h : \text{words} \rightarrow \{0, 1\}^{5 + \lg n} \). If \( s \) is a string of bits, let \( \pi(x, b) \), \( b \in \{0, 1\} \), denote the index of the leftmost bit of \( x \) equal to \( b \)

signature = \( s_1 s_2 \ldots s_5 + \lg n = 00 \ldots 0 \)

for each word \( x \) do \( \text{signature}[\pi(h(x), 1)] := 1 \)

return(\( \pi(\text{signature}, 0) \))

If algorithm returns \( k \), estimate number of distinct words = \( \frac{2^k}{1.5470} \)

If \( \pi(\text{signature}, 0) = 0 \), then leftmost bits of signature = 1110. If there are 16 distinct words on the tape,

\[ \Pr\{\text{none of hash codings begins 0001}\} = \left( \frac{15}{16} \right)^{16} = .356 \]

**Ex:** \( \chi^2 \)-statistic

Given a sample space with \( K \) events (for dice, the \( K = 11 \) events are \( X_i \in \{2, ..., 12\} \)) and a series of \( n \) independent experiments, we expect each event \( i \) to occur \( n \cdot \Pr(X_i) \) times.

\( n = 10000 \) throws of dice

<table>
<thead>
<tr>
<th>Event</th>
<th>( \Pr(X_i) )</th>
<th>( n \cdot \Pr(X_i) )</th>
<th>Observed Number ( Y_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1/36</td>
<td>277.77778</td>
<td>269</td>
</tr>
<tr>
<td>3</td>
<td>1/18</td>
<td>555.55556</td>
<td>520</td>
</tr>
<tr>
<td>4</td>
<td>1/12</td>
<td>833.33335</td>
<td>849</td>
</tr>
<tr>
<td>5</td>
<td>1/9</td>
<td>1111.11111</td>
<td>1088</td>
</tr>
<tr>
<td>6</td>
<td>5/36</td>
<td>1388.88895</td>
<td>1382</td>
</tr>
<tr>
<td>7</td>
<td>1/6</td>
<td>1666.66671</td>
<td>1664</td>
</tr>
<tr>
<td>8</td>
<td>5/36</td>
<td>1388.88895</td>
<td>1429</td>
</tr>
<tr>
<td>9</td>
<td>1/9</td>
<td>1111.11111</td>
<td>1112</td>
</tr>
<tr>
<td>10</td>
<td>1/12</td>
<td>833.33335</td>
<td>836</td>
</tr>
<tr>
<td>11</td>
<td>1/18</td>
<td>555.55556</td>
<td>581</td>
</tr>
<tr>
<td>12</td>
<td>1/36</td>
<td>277.77778</td>
<td>270</td>
</tr>
</tbody>
</table>

Statistic: \( \sum_{1 \leq i \leq K} (Y_i - n \cdot \Pr(X_i))^2 \) For loaded dice, this statistic should be large; fair small. **Problem:** If \( \Pr(X_i) >> \Pr(X_j) \), then probably

\( (Y_i - n \cdot \Pr(X_i))^2 >> (Y_j - n \cdot \Pr(X_j))^2 \).

\( \chi^2 \)-statistic:

\[ \sum_{1 \leq i \leq K} \frac{(Y_i - n \cdot \Pr(X_i))^2}{n \cdot \Pr(X_i)} \]
For above dice data, the $\chi^2$-statistic = 5.917396

\[
\begin{align*}
\text{99%} & \quad 0.00016 & \quad 0.00393 & \quad 0.1015 & \quad 0.4549 & \quad 1.323 & \quad 3.841 & \quad 6.635 \\
\text{95%} & \quad 2.088 & \quad 3.325 & \quad 5.899 & \quad 8.343 & \quad 11.39 & \quad 16.92 & \quad 21.67 \\
\text{75%} & \quad 2.558 & \quad 3.940 & \quad 6.737 & \quad 9.342 & \quad 12.55 & \quad 18.31 & \quad 23.21 \\
\text{50%} & \quad 5.917 & \quad 9.210 & \quad 12.83 & \quad 16.28 & \quad 20.52 & \quad 25.71 & \quad 31.5 \\
\text{25%} & \quad 12.59 & \quad 16.81 & \quad 21.67 & \quad 27.00 & \quad 32.91 & \quad 40.39 & \quad 48.84 \\
\text{5%} & \quad 20.72 & \quad 27.00 & \quad 33.55 & \quad 41.36 & \quad 50.72 & \quad 61.84 & \quad 74.5 \\
\text{1%} & \quad 30.1 & \quad 39.36 & \quad 49.32 & \quad 59.34 & \quad 70.54 & \quad 83.07 & \quad 97.8 \\
\end{align*}
\]

For dice data, there are $K=11$ categories, hence $n=10$ degrees of freedom ($Y_{11}$ depends upon $Y_{1},...,Y_{10}$). The $\chi^2$-statistic should be $>5.917396$ between 95% & 75% of the time, so our deviation is perfectly reasonable (we can not reject the hypothesis that our dice are fair).

**Geometric Distribution**: Let r.v. $X$ denote the number of Bernoulli trials until *success*. $X$ has values in $\{1, 2, 3, ..., \}$. Pr\{$X=k\} = q^{k-1}p$

It’s a legal distribution since Pr\{$X=k\} \geq 0$ for all $k \geq 0$ and

\[
\sum_{k \leq 1} \text{Pr}\{X=k\} = \sum_{k \leq 1} q^{k-1}p = p \sum_{k \leq 1} q^{k-1} = p \frac{1-q}{1-q} = p = 1
\]

\[
E[X] = \sum_{k \leq 1} k q^{k-1}p = p \sum_{k \leq 1} k q^{k-1} = p \sum_{0 \leq k} q^{k-1} = p \sum_{0 \leq k} q^k
\]

**Tangent**: \[
\sum_{0 \leq k} q^k = q + 2q^2 + 3q^3 + ... \quad \text{because} \quad \frac{d}{dq} \left( \frac{1}{1-q} \right) = -1 + 2q + 3q^2 + 4q^3 + ... \quad \text{because} \quad \frac{d}{dx} f(x)g(x) = f(x)g(x) - f'(x)g(x)
\]

\[
\frac{q}{(1-q)^2} = q + 2q^2 + 3q^3 + 4q^4 + ... \quad \text{(end of tangent)}
\]

\[
E[X] = \left( \frac{p}{q} \right) \left( \frac{q}{(1-q)^2} \right) = \frac{1}{p}, \quad \forall (X) = \frac{q}{p^2}
\]

Geometric waiting times are memoryless (and the geometric is the only memoryless distribution). If equipment’s life span is geometric, then it doesn’t age: - waiting time in flipping a coin until heads
- duration of phone conversations within a city
- radioactive decay
- life expectancy of an adult fish
- aging of glass labware

**Theorem**: If a (discrete) process is memoryless, then it’s geometric.

**Proof**: If memoryless, Pr\{$X=k \mid X>k-1\} = \text{Pr}\{X=1\}, \quad 1 \leq k$

By Bayes’s Rule, Pr\{$X=k \mid X>k-1\} = \frac{\text{Pr}\{X=k\}}{\text{Pr}\{X>k-1\}} = \text{Pr}\{X=1\}$

Pr\{$X=k\} = \text{Pr}\{X>k-1\} - \text{Pr}\{X>k\$

\[
(\text{Pr}\{X>k-1\} - \text{Pr}\{X>k\})/\text{Pr}\{X>k-1\} = \text{Pr}\{X=1\}
\]

1 - Pr\{$X=1\} = \text{Pr}\{X>k\}/\text{Pr}\{X>k-1\} = \text{Pr}\{X>k\} | \text{X>k-1} = q \quad \text{for all } k. \quad \text{Â}
Ex: In hashing (to \(b\) buckets) with collision resolution by chaining, how many keys must be inserted until every bucket contains a key? A hit is when a bucket gets its first key, and the \(k\)\textsuperscript{th} stage is the sequence of inserts between the \((k-1)\)\textsuperscript{st} hit until the \(k\)\textsuperscript{th} hit. The first stage consists of the first insert. For the \(k\)\textsuperscript{th} stage, \(Pr(\text{hit}) = \frac{b - k + 1}{b}\). Let \(n_k\) be a r.v. denoting the number of inserts during the \(k\)\textsuperscript{th} stage. The number of inserts for every bucket to get a hit is \(n = \sum_{1 \leq k \leq b} n_k\). Because each \(n_k\) is geometrically distributed (with \(\text{Pr(success)} = \frac{b - k + 1}{b}\)), \(E[n_k] = \frac{b}{b - k + 1}\). By linearity of expectation, \(E[n] = E\left[\sum_{1 \leq k \leq b} n_k\right] = \sum_{1 \leq k \leq b} E[n_k] = \sum_{1 \leq k \leq b} \frac{b}{b - k + 1} = b \sum_{1 \leq k \leq b} \frac{1}{k} = bH_b\).

Poisson Distribution

Let \(X\) be an r.v. denoting the number of events in a given time interval (or region). \(X\) can assume values \(\{0, 1, 2, \ldots\}\).

Let \(p_k(t) = Pr(X = k)\) during time interval \(t\). Assume:

- # events in non-overlapping intervals independent.
- For small \(dt\): \(p_1(dt) = \lambda dt + o(dt)\)

\[\sum_{2 \leq k} p_k(dt) = o(dt)\]

Given independence,

\[p_0(t + dt) = p_0(t)^*p_0(dt)\]

\[p_0(t + dt) = p_0(t)^*[1 - \lambda dt - o(dt)]\]

\[\frac{p_0(t + dt) - p_0(t)}{dt} = - \lambda p_0(t) - \frac{2o(dt)p_0(t)}{dt} = - \lambda p_0(t)\]

Combined with \(p_0(0) = 1\) yields \(p_0(t) = e^{-\lambda t}\)

\[p_1(t + dt) = p_1(t)p_0(dt) + p_0(t)p_1(dt)\]

\[= p_1(t)[1 - p_1(dt)] + p_0(t)p_1(dt)\]

\[p_1(t + dt) - p_1(t) = p_1(t)p_1(dt) + p_0(t)p_1(dt)\]

\[= - p_1(t) \lambda dt + p_0(t) \lambda dt\]

\[\frac{p_1(t + dt) - p_1(t)}{dt} = - \lambda p_1(t) + \lambda p_0(t)\]

\[p_1(t) = (\lambda t) e^{-\lambda t}\]

\[p_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}\]

(poisson distribution)

Legal distribution since:

- \(p_n(t) \geq 0\) for all \(n \geq 0\)

\[- \sum_{0 \leq n} p_n(t) = \sum_{0 \leq n} \frac{(\lambda t)^n}{n!} e^{-\lambda t} = e^{-\lambda t} \sum_{0 \leq n} \frac{(\lambda t)^n}{n!} = e^{-\lambda t} e^{\lambda t} = 1\]
Ex: Flying bomb hits in London during WW2 in regions of area 1/4 km²

\[ \lambda \] \quad \text{t} = 0.9323 \text{ km}²

\begin{array}{cccccc}
& n & 0 & 1 & 2 & 3 & \geq 5 \\
Y_n & 229 & 211 & 93 & 35 & 7 & 1 \\
np_n & 226.74 & 211.39 & 98.54 & 30.62 & 7.14 & 1.57 \\
\end{array}

\[ \chi^2 \text{-} 88\% \] should show worse agreement

Ex: In systems with traffic congestion (queueing), the following are Poisson distributed in a fixed interval:
- # jobs arriving
- # jobs completing service
- # messages through a channel
- # raisins in raisin bread

If \( \Pr\{X = k\} = \frac{e^{-\lambda} \lambda^k}{k!} \), then

\[ \begin{align*}
E[X] &= \sum_{0 \leq k} k \cdot \Pr\{X = k\} = \sum_{0 \leq k} \frac{k e^{-\lambda} \lambda^k}{k!} = \frac{e^{-\lambda}}{k!} \sum_{1 \leq k} k \lambda^k = e^{-\lambda} \sum_{1 \leq k} \frac{\lambda^k}{k!} = \\
&= \lambda e^{-\lambda} \sum_{1 \leq k} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} e^\lambda = \lambda
\end{align*} \]

\[ \begin{align*}
V(X) &= \sum_{0 \leq k} k^2 \Pr\{X = k\} - (E[X])^2 = \sum_{0 \leq k} \frac{k^2 e^{-\lambda} \lambda^k}{k!} - \lambda^2 = \sum_{0 \leq k} \frac{ke^{-\lambda} \lambda^k}{(k-1)!} - \lambda^2 = \\
&= \lambda \sum_{1 \leq k} \frac{ke^{-\lambda} \lambda^{k-1}}{(k-1)!} - \lambda^2 = \lambda \left( \sum_{1 \leq k} \frac{(k-1)e^{-\lambda} \lambda^{k-1}}{(k-1)!} + \sum_{1 \leq k} \frac{e^{-\lambda} \lambda^{k-1}}{(k-1)!} \right) - \lambda^2 = \lambda \left[ \lambda + e^{-\lambda} e^\lambda \right] - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda
\end{align*} \]

Ex: Arrival rate of jobs at computer during interval \( t \) seconds is Poisson distributed with parameter \( 0.3t \).

Pr{3 jobs during 10 sec. interval}?

\[ \frac{e^{-\lambda t} (\lambda t)^3}{3!} = \frac{e^{-3} 3^3}{3!} = 0.22404180765538778 \]

Pr{\leq 20 jobs during 20 sec. interval}?

\[ \sum_{0 \leq k \leq 20} \frac{e^{-\lambda t} (\lambda t)^k}{k!} = \sum_{0 \leq k \leq 20} \frac{6^k}{k!} = 0.9999985448930103 \]

Pr{3\leq k \leq 7 jobs during 5 sec. interval}?

\[ \sum_{3 \leq k \leq 7} \frac{e^{-\lambda t} (\lambda t)^k}{k!} = \sum_{3 \leq k \leq 7} \frac{1.5^k}{k!} = 0.19098360373307222 \]
**Theorem**: If \( X_1, \ldots, X_n \) are mutually independent Poisson distributed r.v.s with parameters \( \lambda_1, \ldots, \lambda_n \), then \( \sum_{k=1}^{n} X_k \) is Poisson distributed with parameter \( \sum_{k=1}^{n} \lambda_k \).

**Normal (Gaussian) Distribution**

**Def**: A (continuous) probability density function \( f \) satisfies:
- \( f(x) \geq 0 \) for all \( x \),
- \( \int_{-\infty}^{+\infty} f(x) \, dx = 1 \)

A cumulative distribution function \( F_X \) of r.v. \( X \) is \( F_X(x) = Pr\{X \leq x\} = \int_{-\infty}^{x} f(x) \, dx \)

where \( f \) is a pdf.

\( F_X \) satisfies:
- \( 0 \leq F_X(x) \leq 1 \), \( -\infty < x < \infty \)
- \( F_X(x) \) is a monotonic nondecreasing function of \( x \)
- \( \lim_{x \to -\infty} F_X(x) = 0 \) and \( \lim_{x \to +\infty} F_X(x) = 1 \)

\( \Pr\{a \leq X \leq b\} = \int_{a}^{b} f(x) \, dx = F_X(b) - F_X(a) \)

**Def**: The normal (Gaussian) probability density function is
\[
 f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2} \quad -\infty < x < \infty \]
The mean is \( \mu \), and the variance is \( \sigma^2 \). A standard normal has \( \mu = 0 \) & \( \sigma = 1 \).

**Central Limit Theorem**: The sum of \( n \) r.v.s (with finite mean & variance) is normally distributed as \( n \rightarrow \infty \), that is,
Let $X_1, \ldots, X_n$ be independent r.v.s, with finite means $E[X_k] = \mu_k$ and variances $V(X_k) = \sigma_k^2$, $1 \leq k \leq n$. As $n \to \infty$, the r.v. $Z_n = \sum_{1 \leq k \leq n} X_k$ is normally distributed with mean $\sum_{1 \leq k \leq n} \mu_k$ and variance $\sum_{1 \leq k \leq n} \sigma_k^2$. 