# Statistical Risk Analysis
# Chapter 2: Probability in Risk Analysis

Shaobo Jin

Department of Mathematics

# Odds

The **odds** for events $A_1$ and $A_2$ are any positive numbers $q_1$ and $q_2$ such that

$$\frac{q_1}{q_2} = \frac{\mathrm{P}(A_1)}{\mathrm{P}(A_2)}.$$

- Knowing probabilities, odds can always be found. But odds do not always give the probabilities of events.
- If $A_1$ and $A_2$ form a partition, then we can find probability as

$$\mathrm{P}(A_1) = \frac{q_1}{q_1 + q_2}, \qquad \mathrm{P}(A_2) = \frac{q_2}{q_1 + q_2}.$$

- Odds is defined up to a scaling constant. For any positive $c$, $cq_1 : cq_2$ is also an odds.

# Compute Odds and Probability

### Theorem (Theorem 2.2)

*Let $A_1$, $A_2$, ..., $A_k$ be a partition of $\mathcal{S}$, having odds $q_i$, i.e. $P(A_j)/P(A_i) = q_j/q_i$. Then,*

$$P(A_i) = \frac{q_i}{q_1 + \cdots + q_k}.$$

### Urn with Balls

Consider an urn with balls of three colours. 50 % of the balls are red, 30 % black, and the remaining balls green. Let $A_1$, $A_2$, and $A_3$ be the ball being red, black, or green, respectively.

1. Find the odds.
2. Compute the probabilities from the odds.

# Bayes' Formula for Odds

- Suppose that two statements $A_i$ and $A_j$ have (*a priori*) odds $q_i^{\text{prior}} : q_j^{\text{prior}}$.
- We know that a statement $B$ about the result of the experiment is true. Then we update our *a priori* odds to *a posteriori* odds, defined by any positive numbers $q_i^{\text{post}}$ and $q_j^{\text{post}}$ such that $q_i^{\text{post}}/q_j^{\text{post}} = \mathrm{P}\left(A_i \mid B\right)/\mathrm{P}\left(A_j \mid B\right)$.
- Bayes' formula implies

$$\frac{q_i^{\text{post}}}{q_j^{\text{post}}} = \frac{\mathrm{P}\left(A_i \mid B\right)}{\mathrm{P}\left(A_j \mid B\right)} = \frac{\mathrm{P}\left(B \mid A_i\right) q_i^{\text{prior}}}{\mathrm{P}\left(B \mid A_j\right) q_j^{\text{prior}}}.$$

Hence, we define

$$q_i^{\text{post}} = \mathrm{P}\left(B \mid A_i\right) q_i^{\text{prior}}, \text{ for all } i.$$

# Example: Risk of Damage

Goods is delivered by lorry, train or airplane. From statistics we know that

| | Lorry | Train | Airplane |
|---|---|---|---|
| Proportion of transportation | 0.50 | 0.30 | 0.20 |
| Proportion of damaged goods | 0.05 | 0.10 | 0.02 |

The prior odds for the modes of transportation are

$$q_L^{\text{prior}} : q_T^{\text{prior}} : q_A^{\text{prior}} = 5 : 3 : 2.$$

Given our data, find the posterior odds of damage for the modes of transportation.

# Update Odds

- Let $q_i^0$ be the a priori odds for $A_i$.
- Let $B_1$, $B_2$, ..., $B_n$, ... be the sequence of statements that become available with time.
- Let $q_i^n$ be the *a posteriori* odds for $A_i$ with knowledge that $B_1$, $B_2$, ..., $B_n$, are true.
- The *a posteriori* odds become

$$q_i^n = \mathrm{P}\left(\text{all } B_1, B_2, ..., B_n \text{ are true} \mid A_i\right) q_i^0.$$

# Recursively Update Odds

$B_1$, $B_2$, ..., and $B_n$ are conditionally independent given $A$ if

$$\mathrm{P}\left(\text{all } B_1, B_2, ..., B_n \text{ are true} \mid A\right) = \prod_{k=1}^{n} \mathrm{P}\left(B_k \mid A\right).$$

Theorem (Theorem 2.3)

*Let $A_1$, $A_2$, ..., $A_n$ be a partition of $\mathcal{S}$, and $B_1$, ..., $B_n$, ... a sequence of true statements (evidences). If the statements $B$ are conditionally independent given $A_i$, then the a posteriori odds after receiving the nth evidence are*

$$q_i^n = P\left(B_n \mid A_i\right) q_i^{n-1}, \quad n = 1, 2, ...,$$

*where $q_i^0$ are the a priori odds.*

# Example: Waste-water Treatment

A chemical analysis of the processed water is done once a day. We are interested in $p = \mathrm{P}(B)$, where $B = \{\text{some standard is satisfied}\}$. By experience from similar stations we claim that the probability $p$ can take values 0.1, 0.3, 0.5, 0.7, and 0.9. We set the prior odds to $q_i^0 = 1$ for all $i$.

Suppose the first 5 measurements resulted in a sequence

$$B \cap B^c \cap B \cap B \cap B.$$

1. Update the odds.
2. Which value of $p$ is the most likely?
3. What is the posterior probability that $p \geq 0.5$?
4. What is the posterior odds of $p \geq 0.5$ versus $p < 0.5$?

# Stream of Events

**Definition (Definition 2.1 and 2.2)**

If an event $A$ is true at times $0 < S_1 < S_2 < \cdots$ and fails otherwise, then the sequence of times $S_i$, $i = 1, 2, \ldots$ will be called a stream of events $A$.

For a stream $A$, let

$$N_A(t) \;=\; \text{number of times } A \text{ occured in the interval } [0, t],$$

and denote the probability of at least one event in $[0, t]$ by $P_t(A) = \mathrm{P}(N_A(t) > 0)$.

Further, for fixed $s$ and $t$, define

$$N_A(s, t) \;=\; \text{number of times } A \text{ occured in the interval } [s, s + t],$$

and $P_{st}(A) = \mathrm{P}(N_A(s, t) > 0)$.

# Initiation Events and Scenarios

- An event $A$ does not necessarily cause hazard for harm or economical losses, e.g., fire ignition.
- In order for $A$ to develop an accident or catastrophe, some other unfortunate circumstances, described by events $B$, have to take place, e.g., faulure of sprinkler system.
- We call $A$ an "initiation event", $B$ a "scenario".
- The probability $P_t(A \cap B)$ is the final goal.
- We treat for simplicity only the case when $B$ can be assumed to be independent of the stream of $A$.
    - We assume that the conditional probability that $B$ is true at time $S_n = s$ does not depend on our knowledge of the stream and whether $B$ occurred or not up to time $s$ and that $S_n = s$.

# Estimate $P_t\left(B \text{ and } A\right)$

- Suppose that $A$ occurs once. Then $\mathrm{P}\left(B \text{ and } A\right) = \mathrm{P}\left(B\right)\mathrm{P}\left(A\right)$ under independence.
- However, $A$ can occur more than once in a time interval of length $t$. Hence, $P_t\left(B \text{ and } A\right) = \mathrm{P}\left(B\right)P_t\left(A\right)$ is usually not correct, even with the independence assumption.
  - But we still use it as an approximation.

# Estimate $P_t(A)$

$$
\begin{aligned}
N_A(t) &= \text{number of times } A \text{ occured in the interval } [0, t] \\
P_t(A) &= \mathrm{P}(N_A(t) > 0)
\end{aligned}
$$

- Let $t$ be one time unit, year say, and define a sequence of random variables $X_i$ as follows

$$
X_i = \begin{cases} 1, & \text{if } A \text{ occured in } i\text{th year}, \\ 0, & \text{otherwise.} \end{cases}
$$

- $\bar{X}$ can be used to estimate $P_t(A)$.

# Estimate $\mathrm{P}(B)$

Estimation of a probability $\mathrm{P}(B)$ can be difficult since $B$ occurs very rarely. Hence $\mathrm{P}(B)$ is often computed by

- mixtures of experts' opinions,
- experiences from similar situations,
- some data of recorded failures of components,
- etc.

For example $\mathrm{P}(B)$ can be taken as a fraction of times when $B$ is true when checked at fixed time points according to some schedule chosen in advance.

# Example: Estimate $\mathrm{P}(B)$

Define

$$
\begin{aligned}
A &= \text{Fire starts,} \\
B &= \text{At least one of the evacuation doors cannot be opened.}
\end{aligned}
$$

We assume that $B$ is independent of the stream of $A$.

- Suppose the safety regulations require periodic tests of functionality of exit doors.
- From the periodic tests, one estimates that on average in 1 per 100 inspections not all the doors could be opened for different reasons, which gives $\mathrm{P}(B) = 0.01$.

# Intensity

Because $A$ can occur multiple times during a period of $t$, $P_t(B \cap A) = \mathrm{P}(B) P_t(A)$ is usually not correct. One possible solution is to choose a small $t$ such that $A$ occurs only once with large probability.

**Definition (Definition 2.3, Intensity)**

For a stationary stream of events $A$, the intensity of events $\lambda_A$ and its inverse $T_A$, called the return period of $A$, are defined as

$$\begin{aligned} \lambda_A &= \lim_{t \to 0} \frac{P_t(A)}{t}, \\ T_A &= \frac{1}{\lambda_A}. \end{aligned}$$

# Some Remarks

1. The concept of stationary is beyond the scope of our course. Intuitively speaking, it means that mechanism creating events is not changing in time.
   - A necessary condition is $P_{st}(A) = P_t(A)$ for any value of $s$.
2. The intensity $\lambda_A$ has a unit.
   - Consider the unit day$^{-1}$:

$$\begin{aligned} \frac{\mathrm{P}(A \text{ occurs in } t \text{ days})}{t} &= \frac{1}{365} \cdot \frac{\mathrm{P}\left(A \text{ occurs in } \frac{t}{365} \text{ years}\right)}{t/365} \\ &= \frac{1}{365} \cdot \frac{\mathrm{P}(A \text{ occurs in } s \text{ years})}{s}. \end{aligned}$$

   - If, for example, $t = 1$ day and $P_t(A) = 10^{-3}$, then

$$\begin{aligned} \lambda_A &\approx 10^{-3} \text{ with unit day}^{-1}, \\ \lambda_A &\approx 0.365 \text{ with unit year}^{-1}. \end{aligned}$$

# Estimation of $\lambda_A$

For a short period of time $t$, we can estimate $P_t(A)$ by $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, $\lambda_A$ can be estimated by $P_t(A)/t$.

- Suppose that $T$ is the time span of our data. Then, it is equivalent to estimating $\lambda_A$ by $N_A(T)/T$, where $N_A(T)$ is the number of times $A$ happened in the time interval $[0, T]$.
- This suggests that

$$\lambda_A = \lim_{T\to\infty} \frac{N_A(T)}{T}.$$

# Independent Streams

**Theorem (Theorem 2.4)**

*Suppose that there are $n$ stationary independent streams where $A_i$ happens with intensity $\lambda_{A_i}$. Let $A$ be an event that any of $A_i$ occurs, i.e. $A = \bigcup_{i=1}^{n} A_i$. Then the stream of $A$ is stationary and its intensity $\lambda_A$ is given by $\lambda_A = \sum_i \lambda_{A_i}$.*

*Consider a scenario $B$ that can happen when $A$ occurs. If $B$ is independent of the stream $A$, then the stream of events when $A$ and $B$ are true simultaneously has intensity*

$$\lambda_{A \cap B} = \lambda_A P(B).$$

# Independent Streams: Consequence

A consequence of $\lambda_A = \sum_i \lambda_{A_i}$ is that even if intensities of accidents $A_i$ are small, it is still likely that some will occur.

- The intensity of fire in a flat $i$ in a building is small.
- However, since there are many buildings in a country, the intensity of fires in any of the buildings in a country can still be high.

# Example: Accidents in Mines

Let $A =$ Accident in a coal mine happens. Our data show that $N_A(40) = 120$, where the time unit is year.

1. Estimate $\lambda_A$ with unit year$^{-1}$.
2. Estimate the probability that $A$ occurs during a time interval of period $t$.

Now let $K$ be the number of deaths in an accident and $B = \{K > 75\}$. Our data show that there are 17 accidents with more than 75 deaths. Suppose that the probability of $B$ is estimated by $17/120$.

1. Estimate $\lambda_{A\cap B}$, if we assume the number of perished $K$ is independent of the stream.

# More Conditions

We have used $P_t(A)/t$ to approximate $\lambda_A$. It is equivalent to $P_t(A) \approx t\lambda_A$. But if $t\lambda_A > 1$, it cannot be used as a probability.

- **C1** More than one event cannot happen simultaneously, i.e. at exactly the same time.
  1. Consider $A =$ An aeroplane crashes. The possibility that two aeroplanes crash at the same instance is negligible and condition C1 holds.
  2. Consider $A =$ A person dies in an aeroplane accident. Condition C1 is not satisfied.
- **C2** The expected number of events observed in any period of time is finite.
- **C3** The number of events that occur in disjoint intervals are independent.

# Poisson Stream of Events

Theorem (Theorem 2.5, Poisson Stream of Events)

*For a stationary stream of event $A$, if conditions C1 and C2 hold, then*

$$P_t(A) \leq t\lambda_A = \frac{t}{T_A},$$

*where $\lambda_A$ is the intensity, and $T_A$ the return period of $A$.*

*If condition C3 also holds, then the number of events $N_A(s,t) \in Po(t\lambda_A)$ (Poisson distribution), viz.*

$$P(N_A(s,t) = n) = \frac{(t\lambda_A)^n}{n!} \exp\{-t\lambda_A\}, \quad n = 0, 1, 2, \ldots$$

*Consequently, the probability of at least one accident in $[0,t]$ is given by*

$$P_t(A) = 1 - P(N_A(t) = 0) = 1 - \exp\{-t\lambda_A\}.$$

# Initiation Events and Scenarios

For a stream $A$ and scenario $B$, suppose that $B$ is independent of the stream and the stream is Poisson.

- From Theorem 2.4, we can show that the intensity is

$$\lambda_{A\cap B} = \lambda_A P_B,$$

  where $P_B = P(B)$ is the probability that $B$ occurs once.

- We can even show that

$$\mathrm{P}\left(N_{A\cap B}(s,t) = n\right) = \frac{\left(t\lambda_A P_B\right)^n}{n!}\exp\left\{-t\lambda_A P_B\right\}.$$

  Hence $A \cap B$ is also a Poisson stream.

# Example: Accidents in Mines

Let $A =$ Accident in a coal mine happens. From historical data, we have obtained

$$\lambda_A \approx 3 \text{ with unit year}^{-1}.$$

We assume that the stream of events is Poisson.

1. Find the probability of more than one accident during the month.

Let $B = \{K > 75\}$, where $K$ is the number of deaths in an accident, assumed to be independent of the stream. Suppose that $P(B) = 17/120$. If $B$ occurs, then we have a catastrophe.

1. Find the probability of at least one serious accident during one month.
2. Find the probability of more than one catastrophe during one month.

# Example: Return Period

Let $A =$ water level exceeds $u^{\text{crt}}$, where $u^{\text{crt}}$ is some critical value.

1. If $t = 1$ year and $P_t(A) = 0.01$, then $u^{\text{crt}}$ is a 100-year water level and $A$ is a 100-year event.
2. For stationary streams, $A$ is a 100-year event if its return period $T_A = 100$ years.

Do these two approaches give different heights for 100-year levels?

# Non-Stationary

Most real phenomenon are non-stationary: environmental conditions vary with time, or systems deteriorate with time.

Definition (Definition 2.4, Intensity, Non-Stationary Case)

Let $s$ be a fixed time point and $P_{st}(A) = \mathrm{P}(N_A(s,t) > 0)$ the probability that at least one event $A$ occurs in the interval $[s, s+t]$, then the limiting value (if it exists)

$$\lambda_A(s) = \lim_{t \to 0} \frac{P_{st}(A)}{t},$$

will be called a non-stationary intensity of the stream of $A$.

If $P_{st}(A)$ does not depend on $s$, then $\lambda_A(s)$ reduces to $\lambda_A$.

# Non-Stationary Poisson Stream of Events

**Theorem (Theorem 2.6, Poisson Stream of Events)**

*Consider a stream of events A. Under some regularity assumptions, if the conditions C1 - C3 are satisfied, then*

$$P_{st}\left(A\right) = 1 - \exp\left\{-\int_{s}^{s+t} \lambda_A\left(x\right)dx\right\},$$

*where* $\lambda_A\left(x\right)$ *is given in definition 2.4.*

*Furthermore, the number of events observed in the time interval* $[s, s+t]$, $N_A\left(s,t\right) \in Po\left(m\right)$, *where* $m = \int_s^{s+t} \lambda_A\left(x\right)dx$.

# Example: Daily Rain

Consider $A_i$ = Daily rain exceeds 50 mm in month $i$, for $i = 1, ..., 12$. We have data of 39 years. Our data show that $N_{A_i}(T)$ are 4, 0, 3, 4, 3, 2, 3, 3, 3, 2, 7, 10.

1. Suppose that January to October have the same intensity, and November and December have the same intensity. Estimate $\lambda_{A_i}$.
2. Suppose that the stream of extreme rains is Poisson with intensity $\lambda_A(s)$. Let $N_1$, $N_2$ be the number of huge rains in the first and second six months during next year, respectively. Find the probability that there will be more than two rains in the periods.